ADDRESSING THE UNCERTAIN GEOGRAPHIC CONTEXT PROBLEM BY ADVANCED SPATIAL ANALYSIS APPROACHES BASED ON GIS AND GPS FOR ENVIRONMENTAL HEALTH STUDIES

BY

JUE WANG

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Geography
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2019

Urbana, Illinois

Doctoral Committee:

Professor Mei-Po Kwan, Chair
Professor Sara McLafferty
Professor Shaowen Wang
Professor Julie Cidell

ABSTRACT

Scholars in the fields of health geography, urban planning, and transportation studies have long attempted to understand the relationships among human movement, environmental context, and their effect on health outcomes. Considerable research has been conducted to advance our understanding of how environmental exposures affects health behaviors and outcomes. In many of these studies, however, environmental exposures are found to have inconsistent associations with health. The uncertain geographic context problem (UGCoP) is one of the most important methodological issues that contribute to the inconsistent findings. This dissertation explores the methodological issues in environmental health research causing the uncertain findings and how the UGCoP influences research findings, proposes an activity space approaches to comprehensively assess the individual exposures to environmental contexts, and designs an innovative environmental exposure evaluation framework to spatiotemporally assess individual environmental exposure and evaluated the environmental effects on health outcomes. With empirical analysis of real-world applications with the GPS tracking data and environmental context data collected at Chicago, IL, and Columbus, OH, these proposed approaches are proved perform better than currently widely used methods. Taking into account the complex spatial and temporal dynamics of individual environmental exposures, the proposed methods also helps to mitigate the UGCoP in important ways. They may be used in environmental health studies concerning environmental influences on a wide range of health behaviors and outcomes, and thus help to improve our understanding of the environmental effects on different health behaviors and outcomes.

Keywords: Environmental health; the uncertain geographic context problem; activity space; environmental context cube; exposure assessment; food environment; physical activity; GPS; GIS

ACKNOWLEDGMENTS

I would like to express my thanks of gratitude to my advisor, Dr. Mei-Po Kwan, for her valuable guidance and unconditional support. She has provided abundant instruction and in-depth comments at every stage of my doctoral study. Without her help and encouragement over the years of pursuing my Ph.D. degree, I would not be able to reach this stage and achieve what I have today.

I much appreciate Dr. Sara McLafferty, Dr. Shaowen Wang, and Julie Cidell for their efforts of serving as the committee. They have provided invaluable suggestions and insights to improve and shape my dissertation. I would also like to express my special thanks to Dr. Marilyn O'Hara Ruiz, who has fully exemplified the high-quality scholarship to which I aspired. In addition, I am indebted to Dr. Yanwei Chai for his help and inspiring discussions in my graduate study.

Finally, I wish to thank my family for their unconditional support and encouragement along the journey of pursuing my Ph.D. degree. I especially appreciate my wife, Qian Ying Lu, for her long-lasting support and endless love, which helps me pass through the difficulties and obstacles on my road of advance. I also own significant gratitude to my parents for their support to my graduate studying abroad. Their everlasting love is my constant source of motivation and encouragement.

To my wife and parents, for their love and support

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION AND BACKGROUND

1.1 MOTIVATION

The World Health Organization reports that about 22% of the global disease burden can be attributed to the environment (Prüss-Ustün *et al.* 2016). A growing body of evidence has also demonstrated an association between health outcomes and environmental exposures. The field of environmental health, which concerns these environmental effects on human health, has attracted increasing levels of attention for its use of advanced geographic information science (GIS) technologies (Macintyre *et al.* 2002). As described by Thacker et al. (1996), environmental health, from a surveillance standpoint, can be categorized into three general areas: hazard surveillance, outcome surveillance, and exposure surveillance. Among these, exposure surveillance, which examines “how people are exposed to environmental hazards and the processes through which exposure results in an adverse health effect” (Maantay and McLafferty 2011), widely uses GIS techniques in exposure analysis (Melnick 2002, Collins 2003) because of their ability to link population data and environmental data (Jerrett *et al.* 2005).

Considerable research has been conducted to advance understandings of the ways in which environmental factors (Handy *et al.* 2002, Saelens, Sallis, Black, *et al.* 2003, Troped *et al.* 2003, 2010, Fisher *et al.* 2004, Mota *et al.* 2005, Hoehner *et al.* 2005, Hillsdon *et al.* 2006, Roemmich *et al.* 2007, Kerr *et al.* 2007, Nagel *et al.* 2008, Grow *et al.* 2008, Maas *et al.* 2008, Sallis *et al.* 2009, 2016, Santos *et al.* 2009, Coombes *et al.* 2010, Mitchell *et al.* 2016, Browning and Lee 2017) influence people’s health behavior and outcomes. Global Positioning System (GPS) tracking data in particular offer a seemingly promising way of evaluating environmental exposure from a more comprehensive perspective (Duncan *et al.* 2009). One systematic literature review of applications of GPS to public health study (Maddison and Ni Mhurchu 2009) concluded that "one major advantage is the ability to collect valuable contextual information, such as the occurrence of activity within specific facilities, and thus improve our understanding about how individuals interact with their environments" (Lachowycz *et al.* 2012). Because GPS tracking data can identify subjects' exact locations at any time, with GIS able to access the environmental context of each location (Krenn *et al.* 2011), the integration of GIS and GPS provides an objective means of examining the correlation between environmental context and health outcomes (Elgethun *et al.* 2003, Wiehe, Hoch, *et al.* 2008, Maddison and Ni Mhurchu 2009).

Many of these studies, however, have been inconsistent in their association of environmental exposures with health (Sallis *et al.* 1998, 2006, King *et al.* 2000, Salmon *et al.* 2000, Trost *et al.* 2002, Hillsdon *et al.* 2006, Maas *et al.* 2008). Recent studies have found three fundamental problems that could contribute to such inconsistency: the modifiable areal unit problem (MAUP) (Lovasi *et al.* 2012, Koohsari *et al.* 2013, Houston 2014), the uncertain geographic context problem (UGCoP) (Dunton *et al.* 2014, Schwanen and Wang 2014, Buck *et al.* 2015, Liao *et al.* 2015), and the spatial non-stationarity problem (Brunsdon *et al.* 1996, Holt 2007, Siordia *et al.* 2012, Wang, Lee, *et al.* 2018). Among these three methodological issues, the UGCoP is one of the most urgent uncertainty-related issues in environmental health studies, for findings of the effects of area-based environmental variables (e.g., land-use mix) on health outcomes or behavior (e.g., physical activity) could be affected by how contextual units or neighborhoods are geographically delineated. The existence of this problem prevents the drawing of any solid conclusions regarding environmental effects on health, with an understanding of the true relationship between environmental exposure and health outcomes needed to inform effective policy interventions that seek to promote public health. Despite ongoing discussions on this methodology uncertainty, scant research has proposed effective methods of addressing the UGCoP. This dissertation investigates this methodological uncertainty issue and proposes innovative methods for assessing individual environmental exposure while demonstrating how methodological innovation can help mitigate the UGCoP.

1.2 RESEARCH QUESTIONS

The overarching goal of this research is to explore the methodological uncertainty in environmental health research and develop a set of novel GIS-based methods for assessing environmental exposure while demonstrating how methodological innovations can help to mitigate the UGCoP. This research is focused on answering the following questions:

1. What methodological issues in environmental health research cause uncertain findings?
2. How can spatiotemporal variations in environmental context be comprehensively profiled?
3. How can individual spatiotemporal environmental exposures be accurately assessed by integrating GPS tracking and GIS?

Answering these pressing questions will require an in-depth interrogation of the relationship between environmental exposure and accessibility, as well as sophisticated improvements to spatial–temporal analysis of GPS tracking and environmental context data.

With a view to addressing these questions, the scope of the dissertation research encompasses the following research objectives:

1. Explore the methodological issues in environmental health research that cause uncertain findings with a view to understanding how the UGCoP problem influences research findings.

2. Develop an innovative method of generating activity space for environmental health research, based on GPS tracking data, such as could comprehensively assess exposure to environmental variables when access to data concerning temporal variations in the environmental context are not available.
3. Establish an approach for spatiotemporally profiling environmental context using GIS data.
4. Design an innovative environmental exposure evaluation framework to spatiotemporally assess individual environmental exposure and evaluate environmental effects on health outcomes.
5. Evaluate these approaches in the context of real-world applications using GPS tracking data and environmental context data collected in both Chicago, Illinois, and Columbus, Ohio.

1.3 BACKGROUND

1.3.1 Environmental Effects on Health Behaviors and Outcomes

Researchers have spent decades examining the relationships between environmental context and health outcomes, during which time there has been an apparent research paradigm shift "from one dominated by a focus on psychological factors and individual responsibility to one recognizing that environmental factors are important in shaping healthy behaviors" (Troped *et al.* 2010). Evidence also indicates that adult mortality is significantly affected by risky health behaviors (Mokdad *et al.* 2004) and that these risky health behaviors are further associated with environmental features (Ross 2000a, Duncan *et al.* 2002, Lee and Cubbin 2002, Cubbin *et al.* 2005).

Environmental effects on health outcomes and behaviors have already been investigated in some depth in previous studies. Abundant research has shown that various health behaviors and outcomes are related to environmental exposure (Millstein *et al.* 2009, Epstein *et al.* 2014, Koohsari *et al.* 2015, Shareck *et al.* 2015). Scholars have found evidence that context exposures influence physical activity (Handy *et al.* 2002, Saelens, Sallis, Black, *et al.* 2003, Fisher *et al.* 2004, Mota *et al.* 2005, Nelson *et al.* 2006, Lee and Moudon 2006, McGinn *et al.* 2007, Berke *et al.* 2007, Carver *et al.* 2008a, Auchincloss 2009, Sallis *et al.* 2009, 2016, Wheeler *et al.* 2010, Ding *et al.* 2011, Almanza *et al.* 2012, Lachowycz *et al.* 2012, Koohsari *et al.* 2015). Various health-related behaviors, including tobacco and drug use, are affected by environmental context exposures (Boardman *et al.* 2001, Xue *et al.* 2007, Kwan *et al.* 2011, Mennis and Mason 2011, Epstein *et al.* 2014, Lipperman-Kreda *et al.* 2015, Shareck *et al.* 2015). Obesity and obesity-related diseases such as type 2 diabetes and cardiovascular disease have also been found to be correlated with environmental exposures (Diez Roux *et al.* 2002, Ellaway *et al.* 2005, Jeffery *et al.* 2006, Nelson *et al.* 2006, Berke *et al.* 2007, Schootman *et al.* 2007, Morenoff *et al.* 2007, Andersen *et al.* 2008, Oliver and Hayes 2008, Seliske *et al.* 2009, Chaix 2009, Millstein *et al.* 2009)—obesity, for example, is more prevalent in areas that lack facilities for physical activity (Giles-Corti *et al.* 2005, Tilt *et al.* 2007) or that are unfriendly to walking (Ewing *et al.* 2003, Frank *et al.* 2004). Furthermore, evidence also indicates that neighborhood context is related to mental health problems (Aneshensel and Sucoff 1996, Ross 2000b, Coker *et al.* 2002, Evans 2003, Fowler *et al.* 2009, Curtis 2010, Stigsdotter *et al.* 2010, Houle and Light 2014, Wheaton and Clarke 2016). Environmental effects on vector-borne disease have also been intensively explored, lest environmental exposure increase the odds of infection (Rogers *et al.* 1996, Acevedo-Garcia 2000, Goetz *et al.* 2000, Lobitz *et al.* 2000, Bavia *et al.* 2001, Anyamba *et al.*

2002, Tatem and Hay 2004). Accordingly, in the field of epidemiology and health geography, understanding environmental exposure is a nontrivial issue involving investigation of environmental effects on human health.

1.3.2 Environmental Exposure Measures

One fundamental question in the field of environmental health is that of how to measure environmental exposure. Although many methods exist, the residential neighborhood is predominantly used as the context unit for measuring environmental exposure, for it is often represented by administrative areas, such as census tracts and postal units, owing to its availability and easy access to routine administrative data. The ready availability of spatial delineations of administrative areas and the lack of detailed mobility data have also contributed to the popularity of administrative areas in environmental health research. With the help of GIS, however, a methodological shift came about in health research, from fixed administration units to ego-centered definitions (Miller 2007, Lee *et al.* 2008, Chaix *et al.* 2009). An ego-centered neighborhood is usually represented by a buffer area centered on an individual's home, with a given threshold of specific distance or travel time (Perchoux *et al.* 2013) that may reflect the exposure area more accurately than administrative units can. However, measuring environmental context with a sole focus on ego-centered neighborhoods may wrongly specify contextual exposure, giving rise to false correlations between environmental context and health outcome (Cummins 2007, Rainham *et al.* 2010). With increasingly broad recognition that neighborhood-based context measures have methodological flaws—because most people spend only a limited amount of time at their home and in its neighborhood—have come calls for more comprehensive evaluation of environmental context that involves consideration of nonresidential locations of daily activities (Kwan and Weber 2003, Inagami *et al.* 2007, Chaix *et al.* 2009, Saarloos *et al.*

2009, Kwan 2012a). Some studies have investigated the correlation between health outcomes and nonresidential environmental context, but most of these have been limited to a specific number of locations (Inagami *et al.* 2006, Jeffery *et al.* 2006, Troped *et al.* 2010, Vallée *et al.* 2010). A better understanding of environmental effects on physical activity requires an accurate and comprehensive assessment of the environmental context of locations where physical activity takes place (Krenn *et al.* 2011).

Amid ongoing debate about the best way of defining geographic context (Weber and Kwan 2003, Inagami *et al.* 2007, Saarloos *et al.* 2009, Kwan 2012a), with the available of activity diary and GPS tracking data, many researchers now believe that the residential neighborhood only partially captures people's exposure to environmental context, with daily activities at other locations also contributing to environmental exposure (Chaix *et al.* 2009, Rainham *et al.* 2010, Houston 2014). The shift from a static measuring approach to a dynamic one has inspired research intended to explore and develop exposure assessment methods with individual activity diary (Arcury *et al.* 2005, Sherman *et al.* 2005, Vallée *et al.* 2010, Chen *et al.* 2011) and GPS tracking data (Duncan *et al.* 2009, Maddison and Ni Mhurchu 2009, Chaix, Méline, Duncan, Jardinier, *et al.* 2013). GPS tracking data, which allow definition of time–location in relation to exposure with unprecedented accuracy (Elgethun *et al.* 2003), are taking environmental exposure assessment from exclusively residential neighborhoods to multiple places of daily activity (Chaix, Méline, Duncan, Merrien, *et al.* 2013). Indeed, GPS has been widely used in other disciplines, such as farming (Schlecht *et al.* 2004), epidemiology (Phillips *et al.* 2001), and transportation (Wolf *et al.* 2003), yet its application for environmental health studies is relatively new (Duncan *et al.* 2009).

The Global Positioning System (GPS) uses satellite data to allow accurate navigation at any point on Earth's surface by providing latitude, longitude, and altitude coordinates with a timestamp. Nowadays, the GPS system is widely integrated with smartphones, smartwatches, and other digital equipment. A portable GPS tracking device can be as small as a matchbox, allowing GPS technology to be easily worn. When study participants wear GPS tracking devices, investigators can track mobility patterns and objectively assess the spatial location of features in the environment or people's behaviors while moving in the environment.

The widespread use of GPS technology in health geography research (Duncan *et al.* 2009, Lachowycz and Jones 2011, Chaix, Méline, Duncan, Merrien, *et al.* 2013, Chaix *et al.* 2016) suggests that GPS tracking data could easily be used to measure varying exposures of different contexts, spatially and temporally, to different people. Portable GPS devices can be used to accurately trace human movement, with advanced GIS methods used to relate these data to high-resolution data regarding relevant environmental risk factors (Almanza *et al.* 2012, Kwan 2012b). A systematic literature review of applications of GPS to physical activity–related health research (Maddison and Ni Mhurchu 2009) concluded that the major advantage of GPS tracking "is the ability to collect valuable contextual information, such as the occurrence of activity within specific facilities, and thus improve our understanding of how individuals interact with their environments and use different locations for physical activity" (Lachowycz *et al.* 2012). GPS tracking datasets provide a promising way of transitioning environmental exposure assessment from exclusively residential neighborhoods to multiple places of daily activity (Zenk *et al.* 2011). Because GPS tracking data can identify subjects' exact locations at any time, and because GIS can access the environmental context of each location (Krenn *et al.* 2011), integration of GIS and GPS provides an objective means of examining the correlation between environmental context and health outcomes (Elgethun *et al.* 2003, Wiehe, Hoch, *et al.* 2008, Maddison and Ni Mhurchu 2009).

What's more, GPS tracking data can be used to generate activity space, an approach that is more representative of people's daily context than residential neighborhoods are (Chaix, Méline, Duncan, Merrien, *et al.* 2013). With accurate tracking of human movement by GPS and integration of context evaluation with GIS, activity space based on GPS tracking data (movement data) appears to be a promising way of assessing the environments that individuals use and to which they are exposed (Krause 2012, Shen and Chai 2013). Activity space comprises "the local areas within which people move or travel in the course of their daily activities" (Albert and Gesler 2003). Because activity space indicates where and how people have contact with their social and physical environments (Golledge 1997), it can be used as a measure of "people's degree of mobility" (Gesler and Meade 1988). The activity space of an individual can thus be used to explore the interaction between human activity and environmental context (Sharp *et al.* 2015, Tamura *et al.* 2017). Standard deviational ellipses, GPS trajectory buffers, minimum convex polygons, and kernel density surfaces have been widely used to represent human activity space (Cummins 2007, Perchoux *et al.* 2013, Sharp *et al.* 2015).

Use of GPS tracking data to delineate activity space is a significant step forward in context exposure assessment. However, Chaix et al. (2013) have argued that health studies that use GPS tracking data may represent a step backward in the drawing of causal inferences between context exposures and health outcomes with respect to the "selective daily mobility bias" in GPS tracking–based context exposure studies, because measures of accessibility to given environmental resources are also determined from the locations that were specifically visited to use the corresponding resources. For example, GPS-based physical activity studies have found that most outdoor physical activity happens in green spaces. It is indeed possible that exposure to green spaces encourages physical activity, but it is also possible that people who want to engage in physical activity prefer to visit green spaces because these places are suitable for physical activity. GPS tracking locations thus reflect actual behavioral contexts rather than potential access, as needed for causal inferences about environmental effects. Furthermore, GPS data cannot capture interactions with context—another crucial factor in understanding the health effects of exposure (Kwan 2012b).

Other limitations of GPS tracking datasets must be carefully addressed when using GPS to analyze the effects of environmental exposures on human health. First, the accuracy of GPS data may vary from location to location—in downtown areas, for example, accuracy tends to be low because of the obstruction and reflection of GPS signals by high-rise buildings (in what is often referred to as the urban canyon effect). Second, GPS can track people in the outdoor environment with high spatial and temporal resolution but cannot collect reliable indoor data, because its signal is blocked by walls and roofs (especially in multilevel buildings and concrete structures). Accordingly, indoor activity and movement patterns cannot be tracked by GPS tracking devices, and other supplemental data relating to indoor activity are thus needed for further analysis of nuanced exposure to context. Third, the sample size of a GPS dataset is normally relatively small owing to the costs of acquiring portable GPS tracking equipment and of recruiting subjects. In addition, although the tracking span of a GPS dataset normally ranges from a week to several months, environmental effects on health outcomes often accrue as part of a long-term (even lifelong) process. Fourth, because GPS tracking datasets contain only location information, the difficulty of deriving the actual activity engaged in at each tracking location may introduce some uncertainty when seeking to measure exposure. For instance, working at a fast-food restaurant versus eating at one might well bring different exposure and effects on body weight. Because context is expected to influence various types of daily activity (e.g., running, reading, eating) differently, supplementing activity data with data on the type and duration of daily activity could provide additional valuable information that promotes better understanding of the role of environmental factors in explaining health behaviors. Fifth, GPS tracking of people may endanger participants' privacy because of the inherent potential for reverse identification. Specifically, analysis of GPS tracking datasets and publication of related results may run counter to individuals' rights to prevent disclosure of the location of their home, workplace, activities, or travel. Because of the high accuracy and individual basis of georeferenced data, geolocations are highly reidentifiable, enabling criminal activity. Privacy concerns thus pose extra challenges for data collection and maintenance of such datasets and prevent the sharing of datasets for multidisciplinary studies.

Rather than using GPS data alone, researchers have tried to integrate other portable sensors to capture both real-time environmental context (e.g., an air pollution sensor to record air quality) and personal status (e.g., an accelerometer to record physical activity level). Measuring the real-time environmental context, for example, portable air pollution sensors can be integrated with GPS tracking to assess individual exposure to air pollution (Greaves *et al.* 2008, Cattaneo *et al.* 2010). Compared to existing methods, which evaluate air pollution by interpolating data from fixed air quality monitoring stations sparsely distributed in space (Gray *et al.* 2013, Nyhan *et al.* 2016), portable air pollution sensors are capable of monitoring and assessing individual exposure accurately in real time. With a view to obtaining real-time personal status, an evaluation of the feasibility of integrating GPS and accelerometer data to assess adult physical activity concluded that data recorded using portable GPS devices were precise enough to allow tracking of individuals' movements sufficient for the purposes of physical activity–related environmental effects research (Rodriguez *et al.* 2005).

Integrating different portable sensors with GPS data could shed light on individual interactions with real-time physical contexts, such as air pollution and noise level, but would not take into account the interplay of social settings and other people. For example, environmental effects (e.g., loud music) may act differently (e.g., depressing or inspiring) on people even in the same context (e.g., home), depending on their immediate mood and the people involved (e.g., close friends or inconsiderable neighbors). Integration of an activity diary with GPS tracking data could enrich those by introducing more detailed information about daily activities while providing new insights into subjects' interactions with social contexts and other people. For example, activity diary data offer information about those with whom an individual interacts, how that individual is feeling, and whether activities are indoor or outdoor—critical pieces of information for assessing context exposure, and ones that cannot be obtained from GPS data alone. Furthermore, the emerging ecological momentary assessment (EMA) offers a promising method for filling in the gap with its ability to capture subjects' real-time behavior, thoughts, and feelings in their immediate environmental context by repeatedly collecting data through handheld electronic devices (Stone and Shiffman 1994, Shiffman *et al.* 2008). The EMA approach has been widely used in clinical psychology and substance addiction research (Stone and Shiffman 1994, Shiffman *et al.* 2002, 2007, Lukasiewicz *et al.* 2007, Epstein *et al.* 2009, Shiffman 2009, aan het Rot *et al.* 2012, Benarous *et al.* 2016). Its feasibility and validity, as demonstrated in previous studies, make it well suited to exploration of environmental effects on health behaviors (Serre *et al.* 2015) from the perspective of contextual individual interaction. However, despite improvements to the theory and methodology used to assess environmental exposure and investigate contextual effects on health outcomes in an activity space, it is expensive and time-consuming for researchers to collect all the kinds of possible data regarding the full spectrum of human activity and substantial methodological challenges remain.

1.3.3 Methodological Problems with Environmental Health Research

Considerable research has been conducted with the aim of advancing current understandings of how environmental factors (Handy *et al.* 2002, Saelens, Sallis, Black, *et al.* 2003, Troped *et al.* 2003, 2010, Fisher *et al.* 2004, Mota *et al.* 2005, Hoehner *et al.* 2005, Hillsdon *et al.* 2006, Roemmich *et al.* 2007, Frank *et al.* 2007, Kerr *et al.* 2007, Nagel *et al.* 2008, Grow *et al.* 2008, Maas *et al.* 2008, Sallis *et al.* 2009, 2016, Santos *et al.* 2009, Coombes *et al.* 2010, Mitchell *et al.* 2016, Browning and Lee 2017) influence people's health behavior and outcomes. Many studies, however, have been inconsistent in their association of environmental determinants with health (Sallis *et al.* 1998, 2006, King *et al.* 2000, Salmon *et al.* 2000, Trost *et al.* 2002, Hillsdon *et al.* 2006, Maas *et al.* 2008). For example, some studies have found no significant association between physical inactivity and weather—a factor commonly perceived as influencing physical inactivity (Sallis *et al.* 1998, 2006, King *et al.* 2000). In the same way, some studies have found access to parks, density of parks within an environment, and proximity to parks to be related to physical inactivity—yet others have not (Cohen *et al.* 2006, Jago, Baranowski, and Baranowski 2006, Jago, Baranowski, and Harris 2006, Roemmich *et al.* 2006, 2007, Frank *et al.* 2007, Tucker *et al.* 2009). Similarly inconsistent results have also been obtained for exposure to recreation facilities (Jago, Baranowski, and Baranowski 2006, Jago, Baranowski, and Harris 2006, Roemmich *et al.* 2006, Timperio *et al.* 2008, Dowda *et al.* 2009, Tucker *et al.* 2009), street connectivity (Braza *et al.* 2004, Roemmich *et al.* 2006, 2007, Frank *et al.* 2007, Kerr *et al.* 2007, 2008, Carver *et al.* 2008b, Larsen *et al.* 2009), and crime-related safety (Jago, Baranowski, and Baranowski 2006, Jago, Baranowski, and Harris 2006, Timperio *et al.* 2008). Three fundamental problems could contribute to such inconsistency: the MAUP (Lovasi *et al.* 2012, Koohsari *et al.* 2013, Houston 2014), the UGCoP (Dunton *et al.* 2014, Schwanen and Wang 2014, Buck *et al.* 2015, Liao *et al.* 2015), and the spatial non-stationarity problem (Brunsdon *et al.* 1996, Holt 2007, Siordia *et al.* 2012, Wang, Lee, *et al.* 2018). Although the MAUP has been widely recognized and intensively studied, the influence of the UGCoP and the spatial non-stationarity problem has been largely ignored and inadequately explored in environmental health studies (Wang, Lee, *et al.* 2018).

1.3.3.1 The modifiable areal unit problem (MAUP)

The MAUP has long been recognized (Zhang and Kukadia 2005, Duncan *et al.* 2013, Wang *et al.* 2014); it results from the use of arbitrary areal divisions (e.g., census tract) and various spatial scales (e.g., neighborhood, city, or county). The MAUP contains two components of potential measurement errors: the scale effect and the zoning effect (Openshaw 1984). The scale effect refers to the inconsistent results caused by the spatial scale used, while the zoning effect refers to the uncertainty prompted by different configurations of zones used for analysis. The MAUP has been intensively investigated in environmental health studies (Mitra and Buliung 2012, Cheng and Adepeju 2014, Clark and Scott 2014, Houston 2014). In previous studies, the results of the analysis have been affected by both the areal units and spatial scale used—for example, the correlations found at various spatial scales could be different (Clark and Scott 2014, Houston 2014).

Kwan (Kwan 2009b) described two perspectives on dealing with the MAUP in environmental health studies. Some scholars seek to identify the best zoning scheme and spatial scale for the specific research problem in the hope of mitigating the MAUP (Openshaw 1996, Mu and Wang 2008). For example, by using homogeneous zones to measure environmental exposures, some measurement errors caused by the MAUP could be mitigated. Others attempt to develop scale-independent analytical measures with which to eliminate the effects of the MAUP (Tobler 1989). Most efforts have sought to build up a scale-independent analytical framework by using the real-time GPS tracking and GIS (Saelens, Sallis, and Frank 2003, Ewing and Cervero 2010, Quigg *et al.* 2010, Wheeler *et al.* 2010, Adams *et al.* 2011, Lachowycz *et al.* 2012, Rainham *et al.* 2012). However, the MAUP still poses a challenge for environmental health studies, and further efforts to handle MAUP properly are needed.

1.3.3.2 The uncertain geographic context problem (UGCOP)

Even use of advanced activity space methods to assess individual environmental exposure has produced inconsistent findings in studies of environmental effects on health behaviors and outcomes (Diez Roux 2001, Oakes *et al.* 2007, Adams and Kapan 2009). Notably, the reliability of existing studies might be affected by misspecification of geographic context (Spielman and Yoo 2009), recently articulated as the UGCoP by Kwan (2012a). The UGCoP refers to the problem whereby findings of the effects of area-based environmental variables (e.g., land-use mix) on health outcomes or behavior (e.g., physical activity) could be affected by how contextual units or neighborhoods are geographically delineated. The problem "arises because of the spatial uncertainty in the actual areas that exert contextual influences on the individuals being studied and the temporal uncertainty in the timing and duration in which individuals experienced these contextual influences" (Kwan 2012b). As discussed by Kwan (2012a), context uncertainty has two critical sources: spatially, the configuration of context units for exposure assessment; temporally, the time and duration of subjects' exposure.

1.3.3.2.1 Spatial uncertainty

As to the spatial configuration of context units, past studies have generally used residential administrative areas, such as census tracts or postal code areas, or buffer areas around individuals' residential locations, as the contextual units (Frank *et al.* 2005, Feng *et al.* 2010, Leal and Chaix 2011, Clark and Scott 2014). Although such units are convenient for assessing context exposure, much research has shown that residential neighborhoods cannot accurately represent the actual areas that exert contextual influences on health outcomes, because people's daily activity covers a much larger area (e.g., whether working, studying, shopping, or engaging in leisure activities) than their residential community (Elgethun *et al.* 2003, Wiehe, Hoch, *et al.* 2008, Kwan 2009b, Basta *et al.* 2010). In addition, because social contexts (e.g., families, friends) do not have predefined geographically defined boundaries, delineating contextual units for these contexts requires consideration of the interactions between social environment and people's daily activities. Furthermore, configuration of contextual units may vary with context and population group. For example, contextual units might be smaller for children and elderly people because of their lower mobility. It is far from clear how to delineate context units, let alone which delineation method appropriately represents true geographic context exposure.

1.3.3.2.2 Temporal uncertainty

People's movement patterns and context variables change dynamically over time. On the one hand, people move about to engage in daily activities, such as work, shopping, and leisure, and they often travel outside their residential neighborhoods to do so. Recent studies of GPS tracking data have provided substantial evidence about where and when people spend their time, indicating that people spend considerable portions of their daily lives outside their residential neighborhood, which has conventionally been defined as the contextual unit for exposure assessment (Cummins 2007, Rainham *et al.* 2010). Day-to-day variability in people's activity locations is also evident (Wiehe, Hoch, *et al.* 2008). Temporal variability of activity locations, of course, may significantly affect the correlations drawn between context exposure and health outcomes. What's more, in addition to variability in daily activities, people also change their residence with time (Kwan 2012a). For long-term effects on health outcomes (e.g., cancer), context exposure assessment should thus consider residential history, including locations lived and the durations lived there.

On the other hand, context variables can vary with space and time. Some context variables change over the twenty-four hours of a day, and others change with the seasons (Gulliver and Briggs 2005, Entwisle 2007). The UGCoP arises if any study seeking to measure context exposure fails to appropriately consider and address spatial and temporal variability in context variables and people's daily activity. Furthermore, time lag between context exposure and health outcomes and the context exposure's cumulative effect could also create failures of association between contextual variables and health outcomes (Chaix, Méline, Duncan, Merrien, *et al.* 2013). The spatial and temporal uncertainties about environmental context and people's movement patterns that are associated with the UGCoP greatly complicate any examination of correlations between environmental context exposures and health outcomes. Failure to identify the actual geographic context thus might lead to inconsistent results and inferential errors.

Existing context exposure assessment methods fail to mitigate the UGCoP both spatially and temporally. From the perspective of spatial uncertainty, existing methods ignore the accessibility of the study area and thus include spaces that may not be accessible to people. What's more, arbitrary cutoff distances are typically used for delineation. Temporally, the duration of environmental exposure is treated merely as a multiplier of exposure, without considering individuals' interactions with space during particular periods. (The more time spent at the location, the more familiar with the surrounding area.)

1.3.3.3 Spatial non-stationarity for environmental health research

Considerable research has been conducted to advance understandings of how physical environmental factors (Handy *et al.* 2002, Saelens, Sallis, Black, *et al.* 2003, Troped *et al.* 2003, 2010, Fisher *et al.* 2004, Mota *et al.* 2005, Hoehner *et al.* 2005, Hillsdon *et al.* 2006, Roemmich *et al.* 2007, Frank *et al.* 2007, Kerr *et al.* 2007, Nagel *et al.* 2008, Grow *et al.* 2008, Maas *et al.* 2008, Sallis *et al.* 2009, 2016, Santos *et al.* 2009, Coombes *et al.* 2010, Browning and Lee 2017) and social environmental factors (Salmon *et al.* 2000, Trost *et al.* 2002, Brown *et al.* 2003, Fontaine *et al.* 2004, Gordon-Larsen *et al.* 2005, Hillsdon *et al.* 2006, Jones *et al.* 2009, Mitchell *et al.* 2016) influence people's level of physical inactivity. Many of these studies, however, have been inconsistent in their association of environmental determinants with physical inactivity (Sallis *et al.* 1998, 2006, King *et al.* 2000, Salmon *et al.* 2000, Trost *et al.* 2002, Hillsdon *et al.* 2006, Maas *et al.* 2008). For example, some studies have found no significant association between physical inactivity and weather, a factor commonly perceived as influencing physical inactivity (Sallis *et al.* 1998, 2006, King *et al.* 2000). In the same way, some studies have identified access to parks, density of parks within an area, and proximity to parks as being related to physical inactivity, whereas others have not (Cohen *et al.* 2006, Jago, Baranowski, and Baranowski 2006, Jago, Baranowski, and Harris 2006, Roemmich *et al.* 2006, 2007, Frank *et al.* 2007, Tucker *et al.* 2009). Similarly inconsistent results have been obtained for exposure to recreation facilities (Jago, Baranowski, and Baranowski 2006, Jago, Baranowski, and Harris 2006, Roemmich *et al.* 2006, Timperio *et al.* 2008, Dowda *et al.* 2009, Tucker *et al.* 2009), street connectivity (Braza *et al.* 2004, Roemmich *et al.* 2006, 2007, Frank *et al.* 2007, Kerr *et al.* 2007, 2008, Carver *et al.* 2008b, Larsen *et al.* 2009), and crime-related safety (Jago, Baranowski, and Baranowski 2006, Jago, Baranowski, and Harris 2006, Timperio *et al.* 2008).

Scholars have attributed such inconsistencies to the MAUP (Lovasi *et al.* 2012, Koohsari *et al.* 2013, Houston 2014) and the UGCoP (Dunton *et al.* 2014, Schwanen and Wang 2014, Buck *et al.* 2015, Liao *et al.* 2015). For example, correlations may differ with spatial scale (e.g., neighborhood, city, county, state) (Houston 2014). What's more, the different delineations of contextual areas (e.g., census tracts, home buffers, road network buffers) used to derive contextual or exposure measures may also affect results (Kwan 2012a, 2012b). Indeed, even at the same spatial scale, using the same exposure assessment method, environmental factors may operate differently at different geographic locations (a phenomenon known as spatial non-stationarity that has been largely ignored in previous studies). For example, street connectivity might be positively correlated with physical inactivity in one city but negatively associated with it in another. To the best of the researchers' knowledge, the spatial non-stationarity of environmental effects on physical inactivity and its influence on the consistency of research findings have not been adequately explored in previous studies and thus are worthy of investigation.

Wang et al. (2018) explored the existence of spatial non-stationarity and illustrated its influence on the research findings by investigating inconsistencies of association between leisure-time physical inactivity (LTPI) and various contextual variables (physical, demographic, and socioeconomic environmental factors). In an exploratory and ecological study, they examined spatial variations of environmental effects on LTPI (percentage of the population reporting insufficient leisure-time physical activity) at the spatial scale of U.S. counties. They found evidence of spatial non-stationarity in environmental health studies. Accordingly, two similar studies of the effects of the same environmental factor on LTPI might indeed give rise to inverse associations if conducted in different regions of the country—perhaps one of the reasons research findings have so often been inconsistent in addressing the effects of environmental factors on health behaviors and outcomes (e.g., (McGinn *et al.* 2007, Kurka *et al.* 2015)). For this reason, assessment of spatial non-stationarity and various environmental effects on LTPI as a function of geographic location may lead to investigations into, and subsequently better understandings of, the ways in which space plays a role in the prevalence of LTPI (Siordia *et al.* 2012).

1.4 THESIS ORGANIZATION

This dissertation comprises three papers that center on the exploration of the UGCoP in environmental health research and on the development of new methods and analysis frameworks for spatiotemporally profiling environmental contexts, assessing individual environmental exposure, and evaluating environmental effects on health outcomes in real-world applications.

Chapter 2 presents hexagon-based adaptive crystal growth Voronoi diagrams based on weighted planes as a way of addressing service area delimitation. It considers the geographic distribution of the clients served by the facilities in question and the characteristics of their socioeconomic context while at the same time mitigating the MAUP when dealing with socioeconomic context by using a method based on continuous weighted planes of socioeconomic characteristics rather than on arbitrary areal units. This study compares a raster grid and a hexagon grid for implementation of adaptive crystal growth Voronoi diagrams, taking as its case study the delimitation of middle school service areas. Its findings indicate that hexagon-based adaptive crystal growth Voronoi diagrams generate better delineation results than the raster-based method, considering the degree to which the population in each service area is commensurate with the enrollment capacity of the middle school in that area and the degree to which middle schools are accessible within their service areas. Although the hexagon-based adaptive crystal growth Voronoi diagrams proposed in this study can help city managers serve their citizens better and allocate public service resources more efficiently, they can also be used in environmental health research to delineate individual activity space based on subjects' daily movement patterns as well as on the environmental context, represented as hexagon-grid weighted planes. Based on the hexagon-based adaptive crystal growth Voronoi diagrams, an innovative context-based crystal-growth activity space method for environmental exposure assessment is developed and discussed in the chapter 3.

Chapter 3 proposes a context-based crystal-growth activity space as an innovative method for generating individual activity space based on both GPS trajectories and environmental context. This method not only considers people's actual daily activity patterns based on GPS tracks but also takes into account the environmental context that either constrains or encourages people's daily activity. Based on GPS trajectory data collected in Chicago, the results indicate that the proposed method generates more reasonable activity space than other existing methods do. The proposed method can comprehensively assess exposure to environmental variables when the data of temporal variations in environmental context are not available, thereby helping mitigate the effects of the UGCoP in environmental health studies.

In Chapter 4, we develop and implement an analytical framework that dynamically represents environmental context (the environmental context cube [ECC]) and effectively integrates individual daily movement (individual space–time tunnel) to allow accurate derivation of individual environmental exposures (the environmental context exposure index). The framework is applied to examine the relationship between food environment exposures and the overweight status of 46 participants using data collected using GPS in Columbus, Ohio, together with binary logistic regression models. The results indicate that the proposed framework generates more reliable measurements of individual food environment exposures than other widely used methods do. Taking into account the complex spatial and temporal dynamics of individual environmental exposures, the proposed framework also helps mitigate the UGCoP. The framework is designed to examine individual food environment exposure but can also be used in a wide range of environmental health studies. It provides an innovative environmental exposure evaluation framework allowing spatiotemporal assessment of individual environmental exposure and evaluation of environmental effects on health outcomes.

Finally, chapter 5 concludes the dissertation with a discussion of the significant findings and limitations of this research as well as of future directions for research.

CHAPTER 2: HEXAGON-BASED ADAPTIVE CRYSTAL GROWTH VORONOI DIAGRAMS[1]

2.1 INTRODUCTION

Delimiting the service areas of public facilities like schools and fire stations is an essential topic in spatial analysis. Geographers and urban planners have a considerable theoretical and practical interest in this topic. The methodology of service area delimitation can be applied to address a wide range of issues as evidenced by the literature on the delineation of market areas (Hess and Samuels 1971, Shanker *et al.* 1975, Marlin 1981, Fleischmann and Paraschis 1988, Sinha and Zoltners 2001, Ríos-Mercado and Fernández 2009), school catchment areas (Yeates 1963, Franklin and Koenigsberg 1973, Holloway *et al.* 1975, Ferland and Guénette 1990, Caro *et al.* 2004), police districts (Benzarti *et al.* n.d., Ma *et al.* 2010, Chen *et al.* 2015, Steiner *et al.* 2015), political districts (Hess *et al.* 1965, George *et al.* 1997, Bozkaya *et al.* 2003, Ricca *et al.* 2013), and service areas of healthcare facilities (Benzarti *et al.* n.d., Ma *et al.* 2010, Chen *et al.* 2015, Steiner *et al.* 2015). There are many existing methods for delimiting service areas, and they can be categorized into mixed-integer programming (Hakimi 1964, 1965, Hess *et al.* 1965, Zoltners and Sinha 1983, Pearce 2000, Williams 2002, Shirabe 2005, Ríos-Mercado and Fernández 2009) and heuristic approaches (e.g., genetic algorithm (Fraley *et al.* 2010), simulated annealing, and tabu search (Openshaw and Rao 1995, Macmillan 2001, Aerts and Heuvelink 2002, D'Amico *et al.* 2002, Bozkaya *et al.* 2003)). Among these methods, Voronoi diagrams, named after Georgy Voronoi, are a popular and promising delimiting method for partitioning space into several subregions based on the distance to a set of seed points specified beforehand (Voronoï 1908).

[1] Reprint, with permission, from Wang, J. and Kwan, M., 2018. Hexagon-Based Adaptive Crystal Growth Voronoi Diagrams Based on Weighted Planes for Service Area Delimitation. ISPRS International Journal of Geo-Information, 7 (7), 257–276.

Voronoi diagrams are widely used in the fields of geography and urban planning because of their computational efficiency and effectiveness for delineating the service area of public facilities (Boyle and Dunn 1991, Boots and South 1997, Okabe and Suzuki 1997, Zhang and Zhou 2004, Zhu *et al.* 2008). Based on the ordinary Voronoi diagram, many advanced Voronoi diagrams have been developed to address the complexity of real-world geographic problems, for instance, the weighted Voronoi diagrams that consider the varying service capacities of public facilities (Schaudt and Drysdale 1991, Ricca *et al.* 2008, Moreno-Regidor *et al.* 2012), the city Voronoi diagrams that take into account the effect of transport networks and travel time (Aichholzer *et al.* 2004), and the adaptive additively weighted Voronoi diagrams that partition a space into subregions with predefined size while considering the position and weight of seed points (Moreno-Regidor *et al.* 2012). More recently, Wang et al. (2014) proposed the adaptive crystal growth Voronoi diagrams based on weighted planes, which consider the geographic distribution of the clients the facilities in question serve and the characteristic of the socioeconomic context. The adaptive crystal growth Voronoi diagram approach has several advantages when compared with other heuristic approaches. It distributes demand over continuous space by weighted planes with fine resolution and, thus, more realistically approximates the spatial distribution of demand. It does not rely on population or demand data based on arbitrary administrative units such as census tracts and thus mitigates the modifiable areal unit problem (MAUP) to a certain extent. The approach takes into account not only the transport network but also walkable areas and natural barriers when evaluating accessibility. The method allows for real-time adaptive growth speed of each service area in order to balance the service load according to their capacity for all facilities.

In spatial optimization studies, the representation issues of demand have been extensively discussed (Miller 1996, Church 1999, Murray 2003, 2008, Murray *et al.* 2008). There are many ways to spatially represent the demand for service, which include point representation (Miller 1996), regional representation (Suzuki and Drezner 1996), and area representation (Love 1972, Bennett and Mirakhor 1974, Aly and Marucheck 1982). Murray et al. (Murray *et al.* 2008) explored the different configurations of coverage modeling and evaluated the effectiveness of various coverage modeling approaches. The representation of demand for service is one of the critical issues for spatial optimization studies and may influence the quality of the spatial delineation results (Murray 2008). For approaches based on an areal representation of demand, there are three kinds of regular polygon grids for covering the land surface without any gap or overlap: equilateral triangles, squares, and hexagons (Carr *et al.* 1992). The equilateral triangular grid is not widely used because the triangles have two different orientations (Birch *et al.* 2007) and complex neighbor relationships. For geographic information science (GIS) modeling and spatial analysis, the square grid is widely used due to its symmetric, orthogonal coordinate system which simplifies the calculation and analysis based on the grid as well as its convenience for transformations between different spatial resolutions (Birch *et al.* 2007). Moreover, the pixels of remotely sensed images are based on square grids (or raster grids), so it is more feasible to utilize square grids to integrate other spatial data with remote sensing data.

The hexagon grid data structure is not widely used in GIS and spatial analysis although it has been utilized in different application fields (e.g., the spatial configuration of cell signal coverage) because of its many merits. Different from the equilateral triangle and square grid structure, the hexagon grid is the most complex regular polygon that can fill a land surface without any gap or overlap (Birch *et al.* 2007). Because hexagonal cells are closer to a circle than rectangular cells in terms of their shape (Feick and Robertson 2015), they are the most compact regular polygons that tile the land surface, so they “quantize the plane with the smallest average error” and “provide the greatest angular resolution” (Sahr *et al.* 2003). Therefore, the hexagon grid can achieve higher accuracy in representing the spatial features of a land surface than can equilateral triangles and square grids from the perspective of spatial analysis (Zook 2015). In addition, from the perspective of visualization and cartography, the boundaries of hexagon cells create less distraction when distinguishing spatial patterns because human eyes have a strong response to the horizontal and vertical lines that are naturally generated by equilateral triangles and square grids (Carr *et al.* 1992). Furthermore, the hexagon grid suffers less from orientation bias, and sampling bias from edge effects since the distances to the centroids of all six neighboring cells are the same. Tessellated hexagon grids are widely used in ecological research, but less utilized in the spatial optimization problem for several possible reasons: (1) The geometry and spatial configuration of the hexagon gird are more complicated than those of the raster grid and, thus, they increase computational burden. (2) Compared with raster data, there is no mature structure for hexagon grid computation that is ready for use by researchers. (3) Remote sensing data, which is one of the most common data resources, is typically stored as pixels in the regular lattice (raster), so utilizing a hexagon grid to process these data involves the conversion from the raster to the hexagon grid data format, which may introduce errors and render data processing more difficult. (4) Processing spatial data with different resolutions is always needed in spatial analysis. Using different resolutions for a raster grid is relatively straightforward because the composition and decomposition of the raster grid maintain congruency and alignment (Birch *et al.* 2007), while it is more complicated for a hexagon grid.

In Wang et al.’s (Wang *et al.* 2014) study, the demand for service and the environmental contexts are rasterized and represented by regular square grids (raster grid). Thus, the purpose of this paper is to compare the raster and hexagon grids for implementing adaptive crystal growth

Voronoi diagrams. Delimitation results of middle school service areas in the study area based on the raster grid and hexagon grid weighted planes will be compared. The objective of the adaptive crystal growth Voronoi diagrams is to achieve the minimized difference of supply/demand ratio among all the facilities while considering the travel time by actual configuration of accessibility-related context features (e.g., road network, walkable area, and natural barriers). Therefore, whether the population in each service area is commensurate with the enrollment capacity of the corresponding middle school and whether the middle schools are accessible within their service areas are examined. The results indicate that the hexagon-grid-based (hereafter hexagon-based) adaptive crystal growth Voronoi diagrams (HACG) produced superior results to the raster-based one considering how commensurate the population in each service area is with the enrollment capacity of the corresponding middle school (supply/demand ratio) and how accessible each middle school is within its service area (travel time) by addressing both the socioeconomic context and accessibility-related context features when delimiting the service areas of schools.

The remainder of the paper is structured as follows. Section 2 introduces the new hexagon-based adaptive crystal growth Voronoi method for delimiting service area based on socioeconomic factors as well as the conventional raster-based method. Furthermore, a case study is presented in Section 3 to delimit the service area of 34 middle schools in the Haizhu district in Guangzhou, China based on the accessibility-weighted and population-weighted planes. In addition, the delimitation results of the hexagon-based method are compared with those of the raster-based method to show the advantage of the hexagon-based adaptive crystal growth Voronoi diagrams. The paper follows with Section 4, which includes a discussion of the advantages and limitations of the hexagon-based method. Section 5 concludes the paper and also suggests some future research directions.

2.2 ADAPTIVE CRYSTAL GROWTH VORONOI DIAGRAMS

This section introduces the adaptive crystal growth Voronoi diagrams based on raster weighted planes as well as the new method based on hexagonal weighted planes. In addition, examples of the raster-based and hexagon-based methods for service area delineation are presented to compare these methods.

2.2.1 Raster-Based Adaptive Crystal Growth Voronoi Diagrams

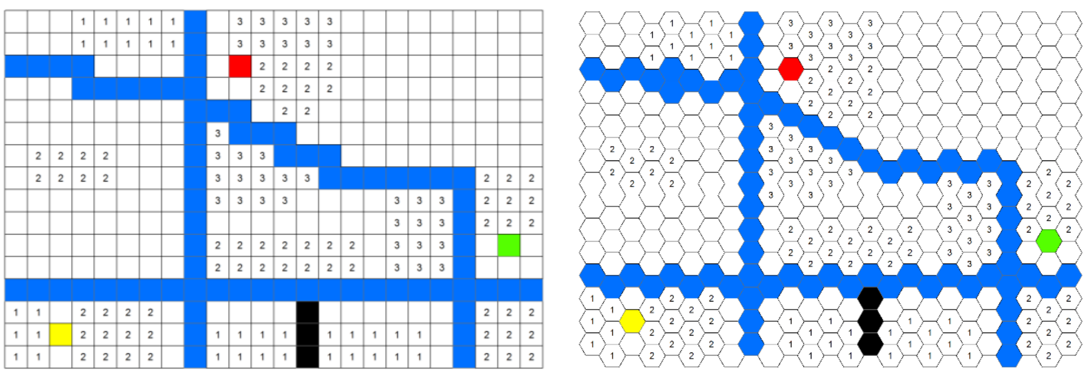
The raster-based adaptive crystal growth Voronoi diagrams (RACG), first proposed by Wang et al. (Wang *et al.* 2014), is an innovative method for delimiting the service areas of public facilities that takes into account the socioeconomic context while mitigating the MAUP. With the middle school service area delineation problem as the case study, the study area was tiled into rectangular cells (raster grid) and two weighted planes (accessibility weighted plane and population distribution weighted plane) were generated based on the raster grid. As illustrated in Figure 2.1a, on the weighted planes, middle schools were represented as seed points (the cells that intersect with the location of middle schools) with their service capacity as the attribute. The accessibility of the research area (e.g., road network, walkable area, and natural barriers) was simulated as different weight values on the accessibility weighted plane, while the distribution of the population was represented as cells with weights on the population distribution weighted plane. In the method, the Voronoi diagrams are grown simultaneously from all seed point cells (e.g., the location of middle schools), and the service area of each middle school will grow to its neighbor cells cycle by cycle. On the one hand, the growth speed at each location (cell) is adjusted based on the attribute of that cell. For instance, roads speed up the growth while rivers or lakes block the growth. On the other hand, the further growth of any grown area can be either encouraged or constrained in real time according to the service capacity of middle schools and the population coverage of their grown areas (supply/demand ratio). For the purpose of illustration, the distribution of the population is simplified and represented as cell values in Figure 2.2 with a range of 1 to 3. As shown in the figure, crystal growth starts from the seed point cells to their neighbor cells, and the growth continues cycle by cycle (Figure 2.2a1 illustrates one crystal growth cycle). Different from other cells, the growth speed of transport network cells is faster due to the higher accessibility they enable (Figures 2.2a1,a2), while growth is constrained by natural barrier cells because it is normally difficult to pass natural barriers such as rivers and lakes (Figures 2.2a2,a3). As the crystal grows, the population coverage of each grown area is calculated in real time at each crystal growth cycle. According to the population coverage of their grown areas, the growth speed is adaptively adjusted based on specific criteria. For example, if the population size of any grown area is 10% larger than any other grown area, the further growth of this grown area will be temporally stopped until its population size is no more than 10% larger than any other grown area in order to balance the population among all grown areas. The crystal growth concludes when all the cells are grown and assigned to the grown area of one seed point (Figure 2.2a4). The delineation results would be the service area of each seed point (e.g., middle school) considering both the distribution of population and the accessibility of the service region.

Note that there are two kinds of neighbor cells on the raster-based weighted planes: 4 orthogonal neighbor cells sharing an edge (or “von Neumann neighborhood”) and 4 diagonal neighbor cells sharing a corner (or “Moore neighborhood”). Because the distance to the centroid of diagonal neighbor cells is larger than that to orthogonal neighbor cells, errors will be introduced if the accessibility to these two types of neighbor cells are not treated differently. Therefore, two region-growing algorithms are introduced for the crystal growth Voronoi diagram: a 4-domain region-growing algorithm that only considers the 4 orthogonal neighbor cells and an 8-domain region-growing algorithm that takes into account both kinds of neighbor cells. Figures 2.3a,b illustrate these two raster-based region-growing algorithms. In Wang et al.'s (Wang *et al.* 2014) method, the 8-domain region-growing algorithm was used for the crystal growth of road network cells due to their high accessibility, while the 4-domain region-growing algorithm was utilized for the crystal growth of other cells.

2.2.2 Hexagon-Based Adaptive Crystal Growth Voronoi Diagrams

Different from the raster grid, which covers the land surface with continuous rectangular cells arranged in columns and rows, the hexagon grid tiles the space with regularly sized hexagonal cells, which are the most complex regular polygons that can fill the land surface without any gap or overlap (Birch *et al.* 2007). Because hexagonal cells are closer in shape to circles than to rectangular cells, they are more compact in space than rectangular cells (Feick and Robertson 2015, Zook 2015). Furthermore, the hexagon grid suffers less from orientation bias, and sampling bias from edge effects since the distances to the centroids of all six neighbor cells are the same (Zook 2015). To the contrary, in a rectangular grid, the distances to the centroids of the 4 neighbor cells that share a corner are longer than the distances to other neighbor cells. Since hexagon cells have only one kind of neighbor cells that share the same edge and have the same distance to their centroids, only one region-growing algorithm is needed for a hexagon grid: a 6-domain region-growing algorithm, which reduces the complication in defining neighbor cells (illustrated in Figure 2.3c).

Similar to raster-based adaptive crystal growth Voronoi diagrams, the distribution of the socioeconomic variable is represented as cell weights in the hexagon grid. For illustration, the socioeconomic context (e.g., population distribution) is simplified and represented in Figure 2.1b with a value range of 1 to 3. Figures 2.2b1–b4 present an example of the hexagon-based adaptive crystal growth Voronoi diagrams based on weighted planes. As illustrated in Figure 2.2b1, the crystal growth starts from the seed point cells based on a 6-domain region-growing algorithm, and the growth continues cycle by cycle. Similar to the raster-based method, the growth speed of transport network cells is faster due to its higher accessibility and growth is constrained by natural barrier cells such as rivers and lakes. The crystal growth speed is adaptively adjusted according to the real-time total weight of each grown area based on the same criteria as for the raster-based adaptive crystal growth Voronoi diagrams illustrated above. As shown in Figure 2.2, the total weight of the grown area in the final result of the hexagon-based method (Red: 87; Green: 88; Yellow: 88) is more balanced than the total weight of the grown area in the raster-based method (Red: 93; Green: 88; Yellow: 82), which indicates that the hexagon-based adaptive crystal growth Voronoi diagrams may work better in service area delimitation problems.



(**a**) (**b**)

Figure 2.1 Illustration of raster-based (**a**) and hexagon-based (**b**) weighted planes. The numbers in the grid cells are the weights representing the spatial distribution of a socioeconomic factor (e.g., population). The blue cells represent elements of the road network; the black cells represent natural barriers; the red, green, and yellow cells are three seed cells representing public service facilities (e.g., middle schools).

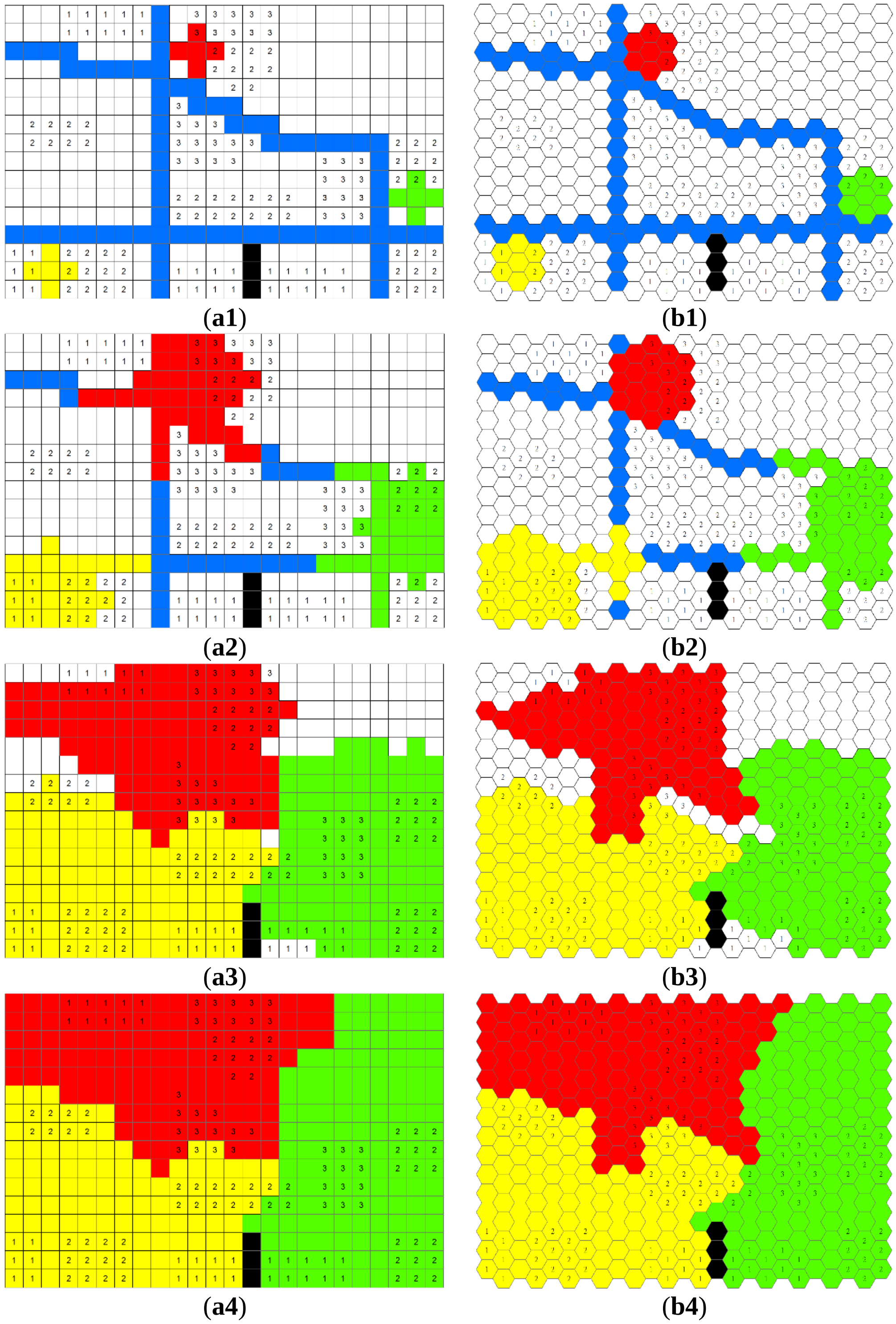


Figure 2.2 Illustration of the adaptive crystal growth Voronoi diagrams with raster-based (**a1**–**a4**) and hexagon-based (**b1**–**b4**) weighted planes. The weights for all grown areas (served population) of the final delineation results are concluded as follows: Red: 93, Green: 88, and Yellow: 82 for the raster-based method; Red: 87, Green: 88, and Yellow: 88 for the hexagon-based method.

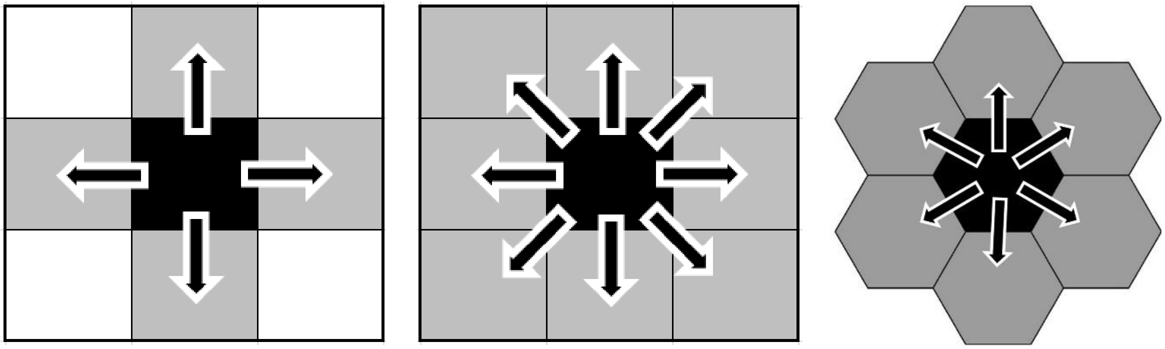


Figure 2.3 The 4-domain (**a**) and 8-domain (**b**) region-growing algorithms for the raster grid; and the 6-domain (**c**) region-growing algorithm for the hexagon grid.

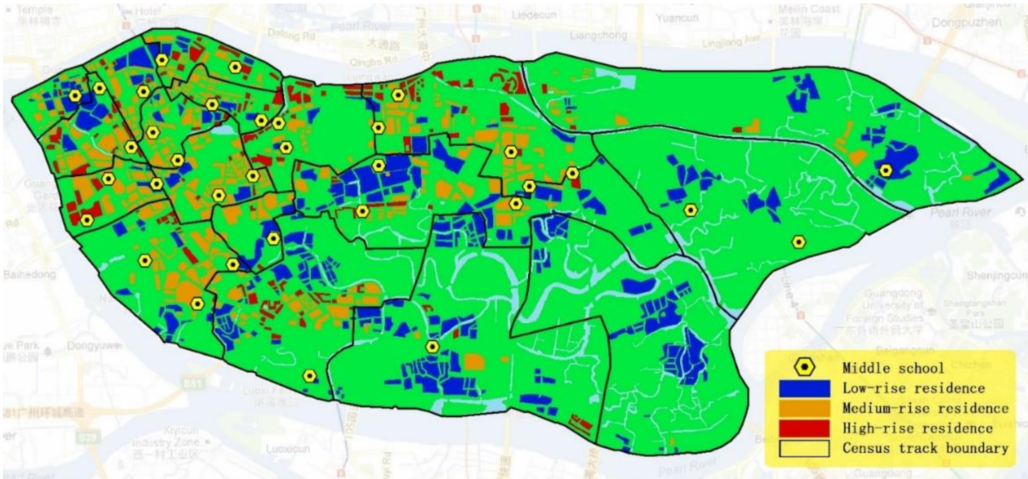
2.3 MIDDLE SCHOOL SERVICE AREA DELIMITATION

In this application, adaptive crystal growth Voronoi diagrams were generated using both raster-based and hexagon-based weighted planes of the same study area with the same socioeconomic context. The results of the HACG will be compared with those of the RACG in delimiting service areas of middle schools using adaptive crystal growth Voronoi diagrams based on accessibility-weighted and population-weighted planes. The delineation results of HACG and RACG are compared in four different spatial resolutions of the weighted plane.

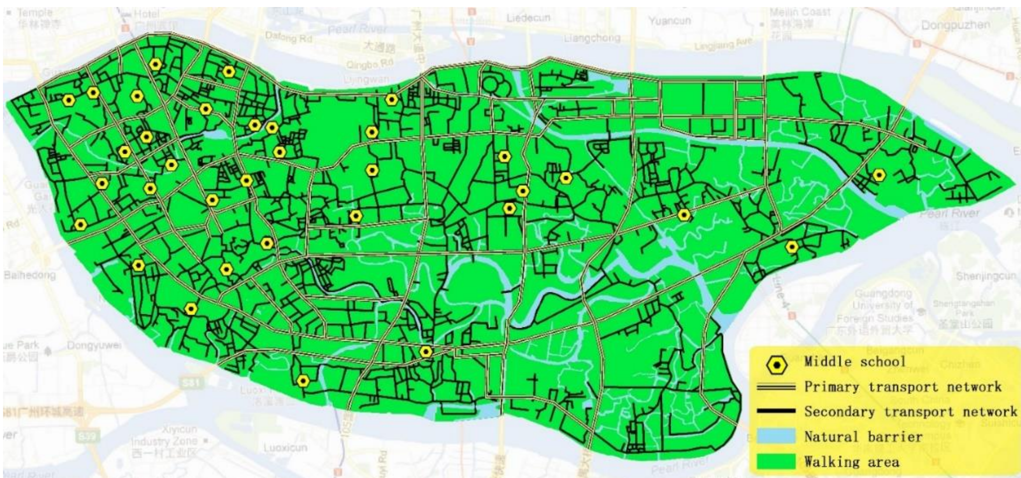
2.3.1 Study Area

Because of the shortage of educational resources, parents in developing countries face key issues for their children's education. Under these circumstances, delimitating the service areas of schools is crucial for ensuring an equal and fair distribution of educational resources among the population. In this study, the Haizhu District in Guangzhou, China was selected as the study area because Guangzhou is one of the largest cities in China that faces the problems of both rapid population growth and shortage in educational resources. The Haizhu District in Guangzhou is an island district (Figure 2.4). It has an area of 90 km^2 and a population of about 1.56 million in 2010 based on the sixth nationwide population census of China. There are 34 middle schools and 18 census tracts in the district. As an island, the Haizhu District is relatively independent of other parts of Guangzhou; therefore, choosing it as the study area would help to reduce boundary effects and edge effects in the analysis. Further, its large population and the considerable number of middle schools make it an ideal location for experimenting with middle school service area delineation.

The delineation of the service areas of middle schools is an allocation problem that spatially allocates demand to service facilities. This allocation problem is suitable to be solved by the adaptive crystal growth Voronoi diagram approach because of the following merits of the method: (1) The adaptive crystal growth Voronoi diagrams distribute the population and environmental characteristics over a continuous surface (covered with rectangular or hexagon grids). They thus eliminate the influence of the boundaries of administrative areas (e.g., census tracts) on delineation results and mitigate the MAUP to a certain extent (although the scale of the grid structure (different grid size) may affect the results). (2) The approach takes into account not only the transport network but also walkable areas and natural barriers when evaluating accessibility. (3) The approach allows for real-time adaptive growth speed of each service area in order to balance the service load according to their capacity for all facilities. This is not possible using other methods. (4) Wang et al.'s (Wang *et al.* 2014) study has justified that the adaptive crystal growth Voronoi diagram approach performs better when compared with other related methods for school service area delineation. That study uses rectangular grids, and the current study extends the method by using a hexagon-grid-based framework and may obtain further improvement in the results.



(a)



(**b**)

Figure 2.4 The location of the study area—Haizhu, Guangzhou (the background map was abstracted from Google Maps)—and the illustration of the population-weighted plane (**a**) and accessibility-weighted plane (**b**).

2.3.2 Weighted Planes

Two types of weighted planes, namely an accessibility-based plane and a population-based plane, were generated for each cell structure (raster grid and hexagon grid). In order to explore the scale effects (how the weighted planes in different scales will affect the delineation results), four different spatial resolutions were used for both raster-based and hexagon-based methods; there are thus one accessibility-based plane and one population plane for each cell structure in each spatial resolution. The accessibility-weighted planes take into account the effect of the transport network and water bodies (as natural barriers) on the accessibility of the schools in the study area. The population-weighted planes represent the distribution of population based on distinct types of residential buildings in the study area. The raster and hexagon cells in these weighted planes have the same size in one spatial resolution to ensure consistency and comparability of the results. Four different spatial resolutions of the raster-based and hexagon-based weighted planes are generated to compare the delineating performance. Specifically, these four spatial resolutions include cell sizes of 50, 100, 150, and 200 m^2.

For the accessibility-weighted planes, illustrated in Figure 2.4a, seed point cells are assigned a cell value of 10 with a growth speed of 1 cell/cycle. The natural barrier cells (e.g., rivers and lakes) are assigned a value of 0 with a growth speed of 0 since they are difficult to pass without a road or bridge. The transport cells are assigned a value of 1 to 2, and the growth speed varies with their average travel speed. The growth speed of the primary transport network (e.g., highways, freeways, and primary trunk roads) cells is 6 pixels/cycle for an average driving speed of about 30 km/h, while the growth speed of the secondary transport network (e.g., secondary trunk roads, collector roads, and local roads) cells is 4 pixels/cycle for an average driving speed of about 20 km/h. For all other cells, including walkable area cells, the growth speed is 1 pixel/cycle for an average walking speed of about 5 km/h. The cell attributes of the accessibility-weighted plane are listed in Table 1.

Table 2.1 Cell values and weights of the accessibility-weighted and population-weighted planes.

	Cell Type	Cell Value	Growth Speed
Accessibility-weighted plane	Natural Barrier Cells	0	0
	Primary Transport Network Cells	1	6 pixels/cycle
	Secondary Transport Network Cells	2	4 pixels/cycle
	Walkable Area Cells	3	1 pixel/cycle
Population-weighted plane	Low-Rise Residential Cells	4	1 pixel/cycle
	Medium-Rise Residential Cells	5	1 pixel/cycle
	High-Rise Residential Cells	6	1 pixel/cycle
	Seed Point Cells (Middle Schools)	10	1 pixel/cycle

The geographic distribution of the population is one of the major considerations for school service area delimitation in addition to accessibility. Thus, another type of weighted plane for the adaptive crystal growth Voronoi diagrams in this study is based on population distribution. As discussed in Wang et al. (Wang *et al.* 2014), due to the modifiable areal unit problem when using census tracts as the zonal scheme, creating better representations of the continuous distribution of the population is crucial for better results in service area delineations. Therefore, the population-weighted planes in this study were created based on census and land use data, which were used to estimate the population distribution of the research area using the same method in Wang et al. (Wang *et al.* 2014). In the method, the continuous population distribution was simulated by distributing the population from census tracts to each cell in the weighted plane based on the distribution and attributes of residential areas. Residential areas are classified according to an urban land use map and urban planning data into three categories: low-rise residential areas, medium-rise residential areas, and high-rise residential areas. With maximum likelihood estimation, population density is 0.0680 persons/m^2 for low-rise residential areas, 0.1135 persons/m^2 for medium-rise residential areas, and 0.0937 persons/m^2 for high-rise residential areas. Because four different spatial resolutions are used in this study, the cell weights of the population were calculated separately by multiplying the population density with the unit cell size. For instance, with the resolution of 100 m^2/cell, the weights of the cells for the three types of residential areas are 6.8 persons/cell (low-rise residential areas), 11.35 persons/cell (medium-rise residential areas), and 9.37 persons/cell (high-rise residential areas) on the weighted plane. Residential cells are assigned a value of 4 (low-rise residence cells), 5 (medium-rise residence cells), or 6 (high-rise residence cells) for the three types of residential areas with a growth speed of 1 cell/cycle. Table

1 lists all the possible cell values and growth speeds on the weighted planes, and Figure 2.4b illustrates the population-weighted plane.

2.3.3 Crystal Growth Configuration

In order to make sure that the results of the raster-based and hexagon-based adaptive crystal growth Voronoi diagrams are comparable, the growth rules for both methods are unified as follows:

(1) The seed cells for the raster-based method are the raster cells that intersect with the locations of the 34 middle schools, while the seed point cells for the hexagon-based method are the hexagon cells that intersect with the locations of the 34 middle schools. In the crystal growth process, the service area of each facility grows from the seed point simultaneously.

(2) In each crystal growth cycle of the raster-based method, the 8-domain region-growing algorithm was used for the crystal growth of road network cells due to their high accessibility, while the 4-domain region-growing algorithm was utilized for the crystal growth of other cells. Regarding the hexagon-based method, the 6-domain region-growing algorithm was used.

(3) If a neighbor cell has already been labeled as a grown area or it is a barrier cell, it will be skipped. Otherwise, the neighbor cell will be labeled as a grown area of the specific middle school. Whenever a particular cell is incorporated into the grown service area of a particular seed, it becomes part of that service area. In other words, the calculation is based on the rules of "first come, first served." If the growth patterns of two service areas come into contact with each other at one place, the growth at that border location will stop for both service areas because the cells beyond that point are already incorporated into either of the service areas. If a cell is claimed by two or more different service areas at the same cycle during the crystal growth process, this cell will be randomly assigned to one of the service areas.

(4) The speed of crystal growth for each cell is determined by the cell value and its growth speed, which was defined on the accessibility-weighted plane. Therefore, the speed is 1 cell per crystal growth cycle for seed point cells, walkable area cells, and residential area cells, while the speed is 4 cells per cycle for secondary transport network cells (the average travel speed on secondary transport network is about 4 times the speed of walking) and 6 cells per cycle for primary transport network cells (the average travel speed on primary transport network is about 6 times the speed of walking).

(5) The growth speed of each grown area is adaptively adjusted in real time to optimize the population in each service area because the population in each service area should be commensurate with the enrollment capacity of each middle school. Therefore, the enrollment capacity of middle schools and the corresponding proportions of enrollment capacity out of the total enrollment capacity in the study area are considered as benchmarks for adjusting the crystal growth speed of each service area. The population of each grown area is calculated based on the population-weighted plane in real time. If the proportion of the population of a grown area (P_i) is larger than the proportion of its enrollment capacity (PEC_i) by a defined value ω (e.g., ω = 10%,), the growth of the corresponding middle school will be restricted:

$$PEC_i = \frac{EC_i}{\sum_{k=1}^{N} SC_k} * 100\% \; (1 \leq i \leq N) \tag{2.1}$$

$$AW = \sum_{k=1}^{N} W_k \tag{2.2}$$

$$P_i = \frac{W_i}{AW} * 100\% \; (1 \leq i \leq N) \tag{2.3}$$

where N is the number of middle schools, EC_i is the enrollment capacity of middle schools, PEC_i is the proportion of enrollment capacity of each middle school out of the total enrollment capacity in the study area, W_k is the population in each grown area, and P_i is the proportion of the population of each grown area out of the total population in all the grown areas.

If $P_i > PEC_i * (1 + \omega)\ (1 \le i \le N)$, then the growth of middle school i will be restricted until $P_i \le PEC_i * (1 + \omega)\ (1 \le i \le N)$. ω is the parameter that defines the strictness of the crystal growth constraint, and it is calibrated to achieve the best result. Generally speaking, the lower the ω value, the stricter the growth control; the higher the ω value, the less restrictive the crystal growth. For the specific case study of middle school service areas, the capacity of the service area of each middle school is controlled in the following manner. Equations (1)–(3) are used to calculate the proportion of the population of a grown area (P_i) and the proportion of its enrollment capacity (PEC_i) among all middle schools in the study area. Both values are calculated for each grown area after each growth cycle. For any grown service area, if the P_i is larger the PEC_i by a defined value ω (e.g., ω = 10%), the growth of the corresponding middle school will be restricted. In the calculation, the ω value will iterate from 0.001 to 0.1 in steps of 0.005 to find the best setting for the delineation considering the supply/demand ratio of the results.

(6) The crystal growth will finish if all the accessible cells are grown and assigned to the grown area of one seed point (middle school).

ArcMap was used for the data preprocessing and the hexagon grid generation. The simulation of the service area delineation was implemented with an algorithm we developed based on Python 2.7 and the open source package of Geospatial Data Abstraction Library (GDAL/OGR Contributors 2018). The calculation was performed on a Windows 10 desktop computer (CPU: Intel Core i7-6700K 4 GHz; RAM: 16 GB).

2.3.4 Hexagon Data Structure

The raster-based weighted planes for RACG are generated based on the widely utilized raster data structure. Nevertheless, there is no mature data structure for hexagon grid computation to the best of our knowledge. The hexagon tessellation implemented by ArcGIS is in the vector format (collection of polygon features), which leads to very low computational efficiency. To address the limitation, we designed a hexagon data structure (we named it "Hexter") in a similar format to a raster grid but specifically for the crystal growth algorithm to increase the computational efficiency in this study. As illustrated in Figure 2.5a, the original hexagon grid is indexed as "ColID-RowID", in which "ColID" represents the index of the column (ranging from "A" to "ZZZ") in the grid, while "RowID" indicates the index of the row (ranging from "1" to "*N*") in the grid. With the indexing, the hexagon grid in the vector format is converted to a raster-like data structure (see the Hexter table in Figure 2.5b) with their "ColID" as the columns and "RowID" as the rows in the Hexter table. With the change of the data format, the spatial relationship in the table is also changed. Since the only spatial relationship used in the crystal growth algorithm is the neighbor cells, we defined the neighboring relationship of the Hexter based on the original relationship of the hexagon grid. Illustrated in Figure 2.5b, for the cells in the odd columns (e.g., C, E, and G), the neighbors are the cells in their top, bottom, left, right, bottom-left, and bottom-right (e.g., cells with indices "C-2" and "C-7" in Figure 2.5), while for the cells in the even columns (e.g., B, D, and F), the neighbors are the cells in their top, bottom, left, right, top-left, and top-right (e.g., cells with indices "H-2" and "H-7" in Figure 2.5). The rows do not influence the neighboring relationship of cells. Note that the Hexter table does not represent the spatial relationship of hexagon cells. It is only used to read the vector-formatted hexagon grid into the memory in the format of a table for crystal growth computation. After the computation, the crystal growth results (cell values) will be projected to the original hexagon grid by matching the "ColID-RowID".

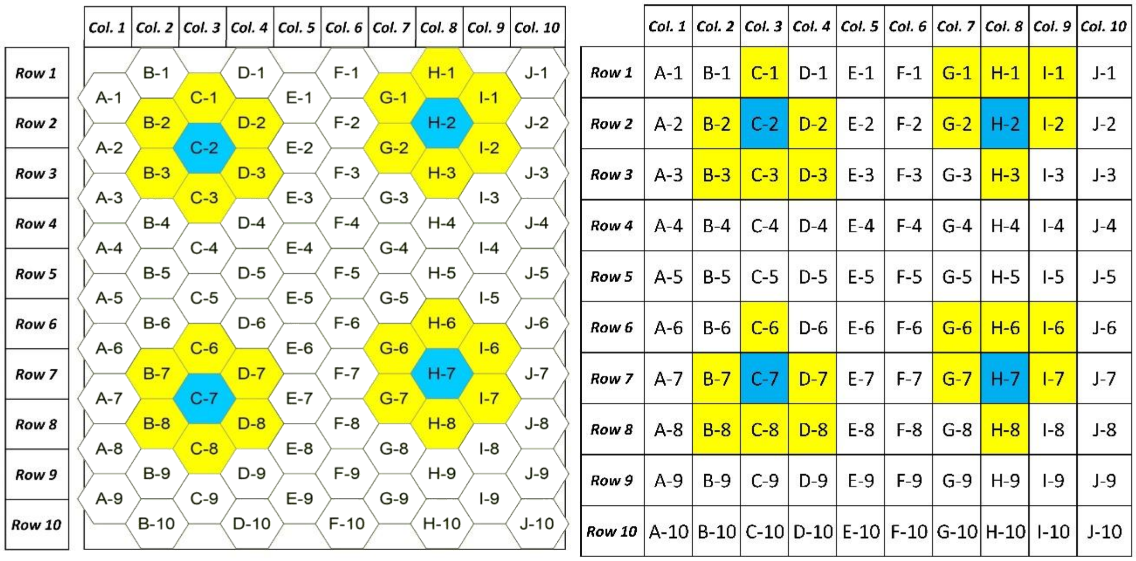


Figure 2.5 The original hexagon grid in the vector format (**a**) and transformed hexagon grid in the "Hexter" data structure (**b**) with an illustration of the neighbor cells definition. The illustrated center cells are colored in blue while their immediate neighbor cells are colored in yellow.

2.3.5 Results

In the study, middle school service areas were delimited by RACG and HACG in four different spatial resolutions (R1: 50 m^2; R2: 100 m^2; R3: 150 m^2; R4: 200 m^2) with different ω settings ranging from 0.001 to 0.1. In order to compare the results, two types of measurements are utilized: (1) how commensurate the population in each service area is with the enrollment capacity of the corresponding middle school; and (2) how accessible each middle school is within its service area.

The root-mean-square error (RMSE) and maximum error (Max. Error) are employed to compare the commensurateness between the enrollment capacity of a school and the population of its service area. Figure 2.6 shows the RMSE of the differences between the proportion of enrollment capacity and the proportion of the population in the service areas delimited by RACG and HACG. It can be seen from the figure that, with different spatial resolutions, the HACG (solid lines in the figure) generally has smaller RMSE compared with the RACG (dashed lines in the figure). In addition, the finer the spatial resolution, the better the results considering the RMSE. In the figure, all lines present a U shape. The best results for different methods and different resolutions are generally achieved when ω has a value between 0.015 and 0.065. If ω is larger or smaller than these values, the RMSE increases. For the HACG, the best observed result is achieved when the spatial resolution is 50 m^2 and ω is 0.055 or ranging from 0.015 to 0.035. Similarly, the best observed result is achieved when the spatial resolution is 50 m^2 and ω is between 0.015 and 0.035 for the RACG. However, comparing the two curves of HACG and RACG with the spatial resolution of 50 m^2, the delimitation results of the hexagon-based method generally perform better with lower values of RMSE. Figure 2.7 illustrates the Max. Error of the differences between the proportion of enrollment capacity and the proportion of the population in the service areas delimited by the RACG and HACG with different resolution and ω settings. Similar with the results of RMSE, the HACG (solid lines in the figure) generally has smaller Max. Error compared with the RACG (dashed lines in the figure) with different spatial resolutions. For both RACG and HACG, the finer the spatial resolution, the smaller the Max. Error. According to the figure, the best observed results for RACG are achieved when ω is ranging from 0.001 to 0.015, while the best observed results for HACG are achieved when ω is ranging from 0.02 to 0.08. Considering both the RMSE and Max. Error, the best observed result is obtained when ω is ranging from 0.02 to 0.035 for HACG, while it is achieved when ω is 0.015 for RACG. With the best observed ω values, the HACG has much lower RMSE and Max. Errors compared with the RACG, as shown in the figure. According to the results, both the Max. Errors and RMSE of the differences between the proportion of enrollment capacity and the proportion of population in the service areas are lowest when using the hexagon-based method after calibration of the parameter ω. Therefore, the results indicate that the hexagon-based method performs better considering the enrollment capacity of middle schools and the population in their service areas.

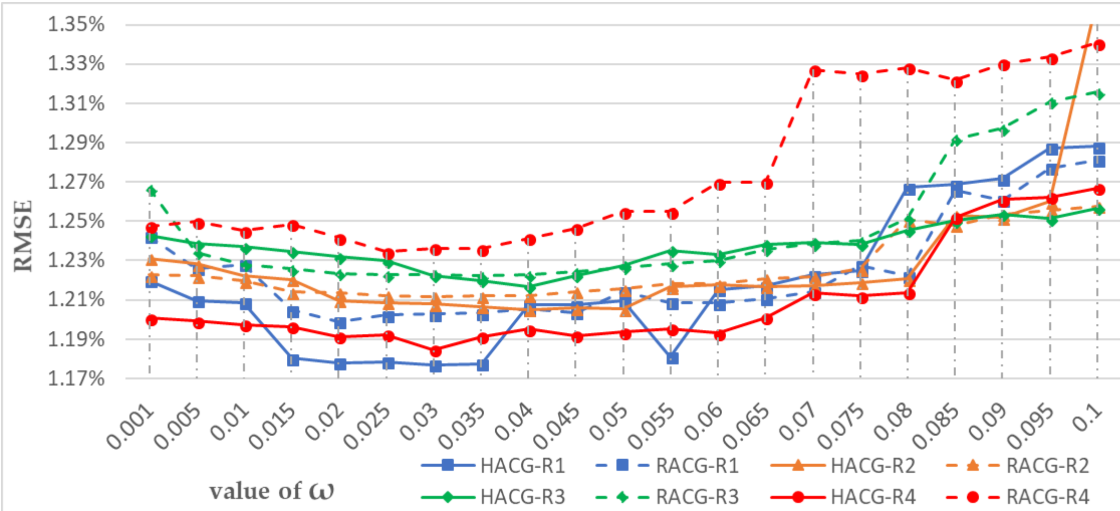


Figure 2.6 The root-mean-square error (RMSE) of the differences between the proportion of enrollment capacity and the population proportion of service areas with different ω settings. RACG: raster-based adaptive crystal growth method. HACG: hexagon-based adaptive crystal growth method.

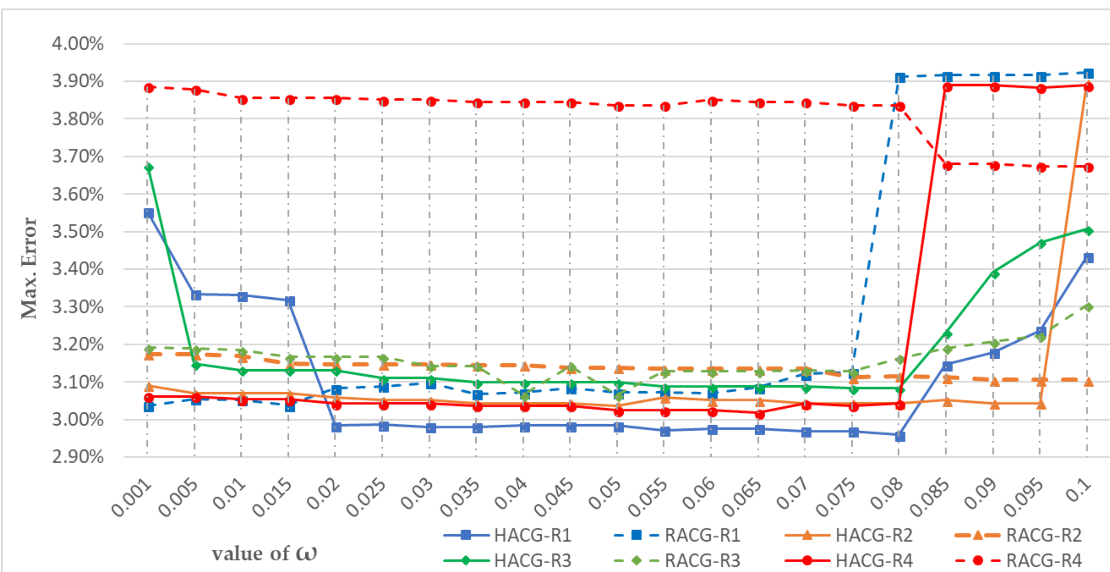


Figure 2.7 The maximum error of the differences between the proportion of enrollment capacity and the population proportion of service areas with different ω settings.

The longest travel time (LTT) to a middle school from any location in its service area is used as a measure of the accessibility to the middle school within its service area in this study. The maximum LTT indicates the lowest accessibility of the middle school in its service area. As discussed above, the best observed results are achieved for both RACG and HACG when the spatial resolution is 50 m^2 considering the enrollment capacity of middle schools and the population in their service areas; therefore, the accessibility to the middle school within its service area of the delineation results is only compared using a spatial resolution of 50 m^2. Figure 2.8 shows the maximum LTT of the service areas of all middle schools in the study area based on the delimitation results obtained with different ω settings. For the HACG, the maximum LTT varies with different ω settings: it stays at high values of around 24 min when $0.035 \leq \omega \leq 0.05$, while it is as low as 21 min with other ω values. Regarding the RACG, the lower values of maximum LTT (around 25 min) are achieved when ω is smaller than 0.02 and larger than 0.045. Comparing the two curve lines of RACG and HACG, it is clear that HACG always performs better than RACG with smaller maximum LTT.

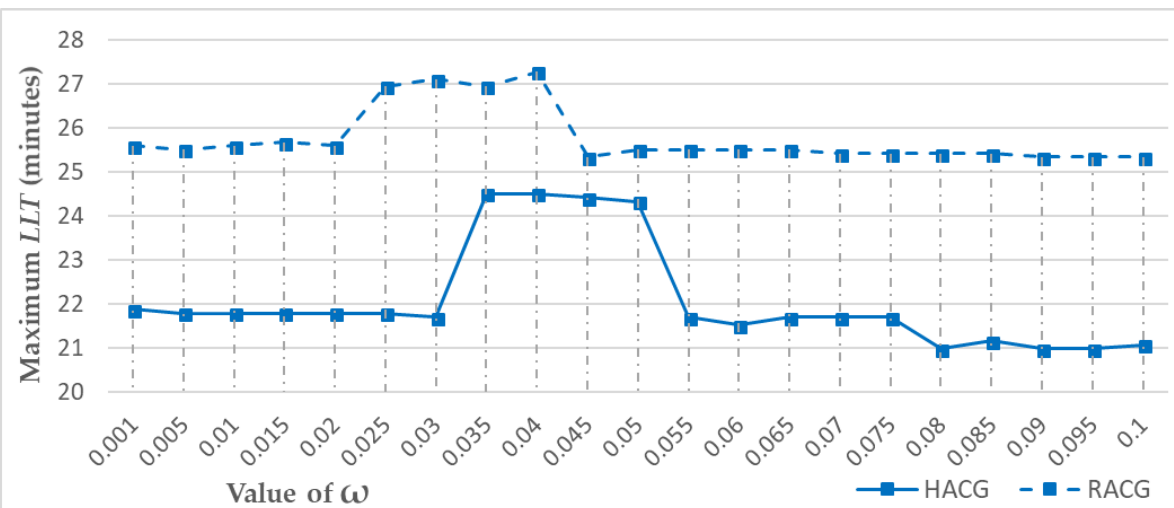


Figure 2.8 Maximum longest travel time (LTT) of service areas based on the delimiting results with different ω settings.

The mean LTT is the average value of the longest travel time to the middle schools within their service areas, which is used as a general indicator of the accessibility of all middle schools in their service areas based on the delimitation results. Figure 2.9 illustrates the mean LTT of the service areas of all middle schools based on the delimitation results obtained with different ω settings. The figure shows that the mean LTT for the HACG stays around 8.4 min while it is around 9.4 min for the RACG, but the mean LTT slightly increased with the increase of the value of ω for both methods. Like the maximum LTT, the mean LTT obtained with the RACG is always higher than that obtained with the HACG.

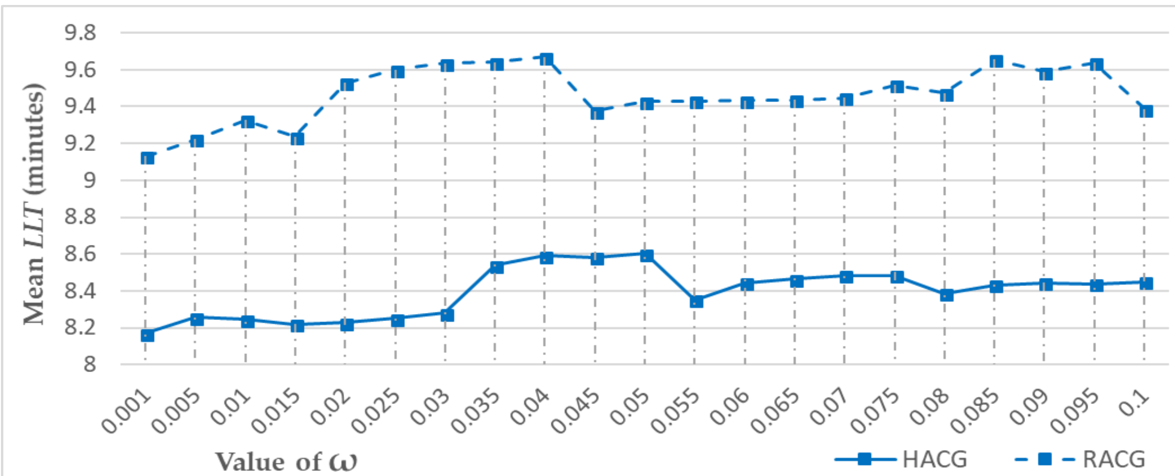


Figure 2.9 Mean LTT of service areas based on the delimiting results with different ω settings.

The standard deviation of LTT measures the amount of variation of the longest travel time to the middle schools within their delineated service areas. Figure 2.10 illustrates the standard deviation of the LTT for all middle schools based on the delineated service areas with different ω settings. Similar to the maximum LTT, for the HACG, the standard deviation of LTT stays at high values of around 4.9 min when $0.035 \leq \omega \leq 0.05$. For all the other ω settings, it stays at low values of around 4.2 min. On the contrary, the standard deviation of the LTT obtained with the RACG is always larger than 5 min. Again, the standard deviation of LTT obtained using the RACG is always higher than that obtained with the HACG.

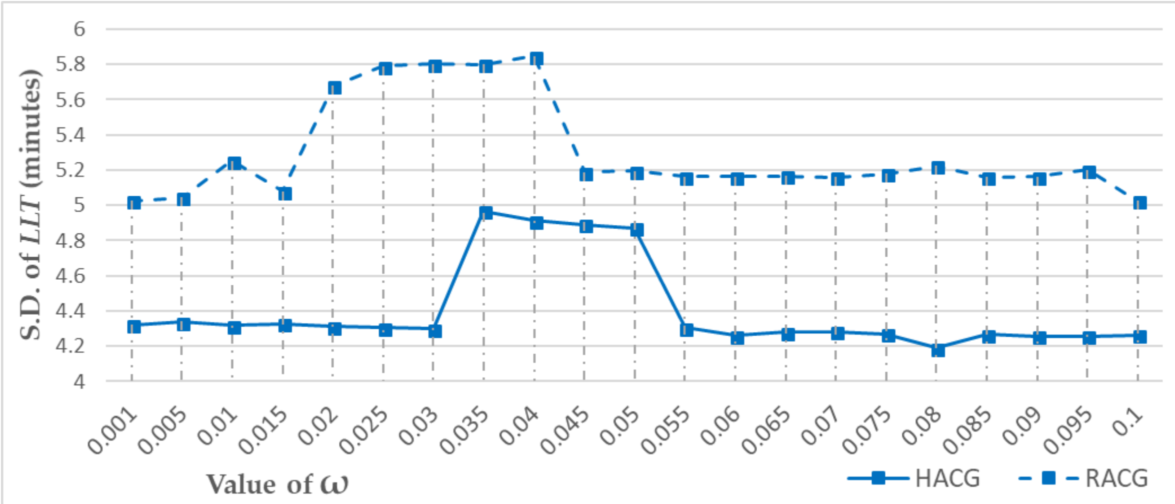
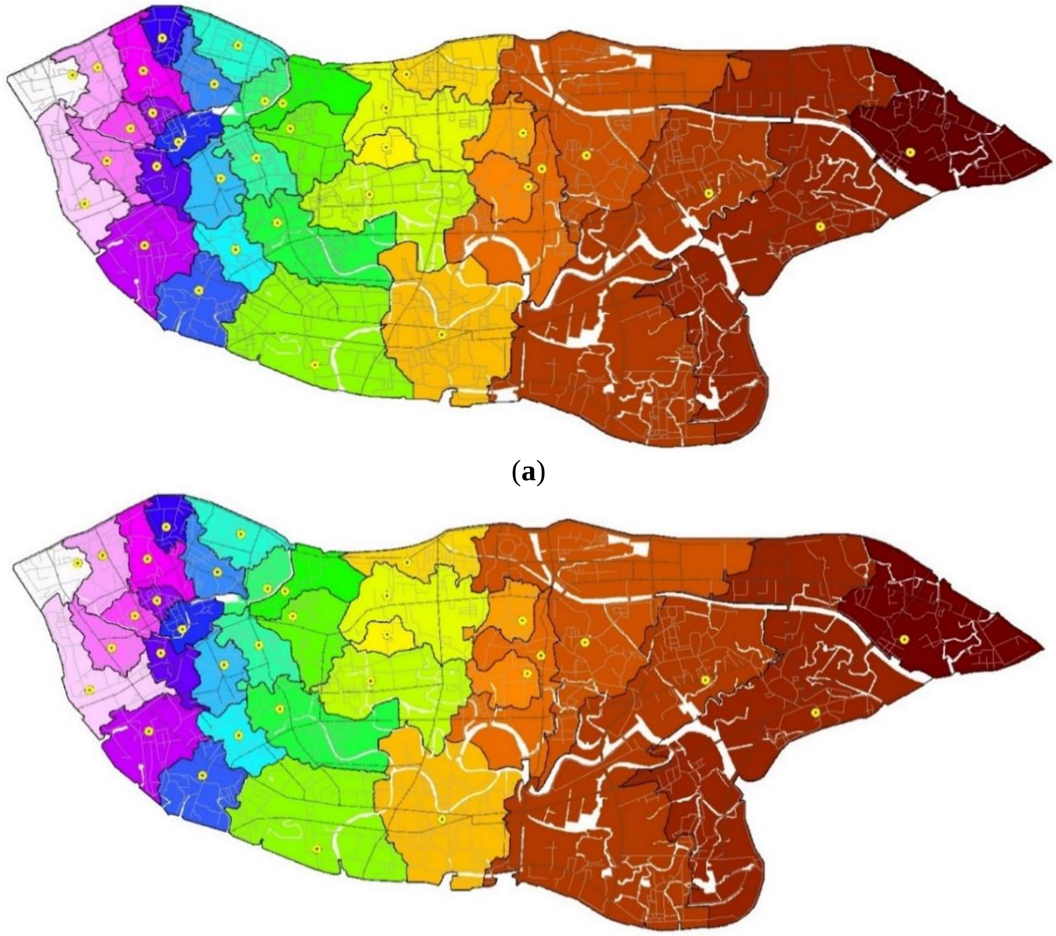


Figure 2.10 The standard deviation of *LTT* of service areas for all middle schools based on the delimiting results with different ω settings.

According to the calibration results of the ω settings, the best delimitation results were selected based on how commensurate the population in each service area is with the enrollment capacity of the middle school in the service area and how accessible the middle schools are within their service areas. Thus, the observed best value of ω for the HACG is 0.02, while it is 0.015 for the RACG. Figure 2.11 illustrates the service area delineation results of the adaptive crystal growth Voronoi diagrams using the raster-based and hexagon-based weighted planes with their respective best observed *ω* values. It can be seen from the figure that the results are generally different.



(**a**)

(**b**)

Figure 2.11 The delineation result of middle school service areas by using raster-based (**a**) and hexagon-based (**b**) adaptive crystal growth Voronoi diagrams.

Further, the RMSE and Max. Error of the differences between the proportion of enrollment capacity and the proportion of the population in the service areas, as well as the maximum, mean, and standard deviation of the LTT to a middle school from any location in its service, are compared in Table 2 with the observed best results of both methods. As the table shows, the RMSE, Max. Error, Max. LTT, mean LTT and S.D. LTT are all smaller when using the hexagon-based method compared with their values obtained when using the raster-based method, which also indicates that the HACG performs better than the RACG considering both the commensurate the population in each service area is with the enrollment capacity of the corresponding middle school and the accessibility of each middle school within its service area.

Table 2.2 Comparison of the differences between the proportion of enrollment capacity of each middle school and the proportion of the population in each service area of the 34 middle schools in the observed best delimitation results (ω = 0.015 for RACG; ω = 0.02 for HACG).

	RMSE	**Max. Error**	**Max. *LTT***	**Mean *LTT***	**S.D. *LTT***
RACG	1.2048%	3.0379%	25.6759	9.2379	5.0803
HACG	1.1779%	2.9833%	21.7985	8.2283	4.3100

Note: RACG represents the raster-based adaptive crystal growth Voronoi diagram; HACG represents the hexagon-based adaptive crystal growth Voronoi diagram.

Considering calculation efficiency, Figure 2.12 illustrates the number of cells for the zoning with different spatial resolutions and their computation time for both raster-based and hexagon-based adaptive crystal growth Voronoi diagrams. It can be seen from the figure that the number of cells increases (from 0.7 to 2.8 million) with finer resolution (from 200 to 50 m^2), and the computation time for both methods rises. The computation time for RACG increases from 0.09 to 0.3 minutes while that for HACG increases from 0.35 to 2.34 minutes. With the longest computation time of about 2 minutes with a personal computer, the calculation efficiency for HACG is acceptable and feasible to be utilized in service area delineation problems.

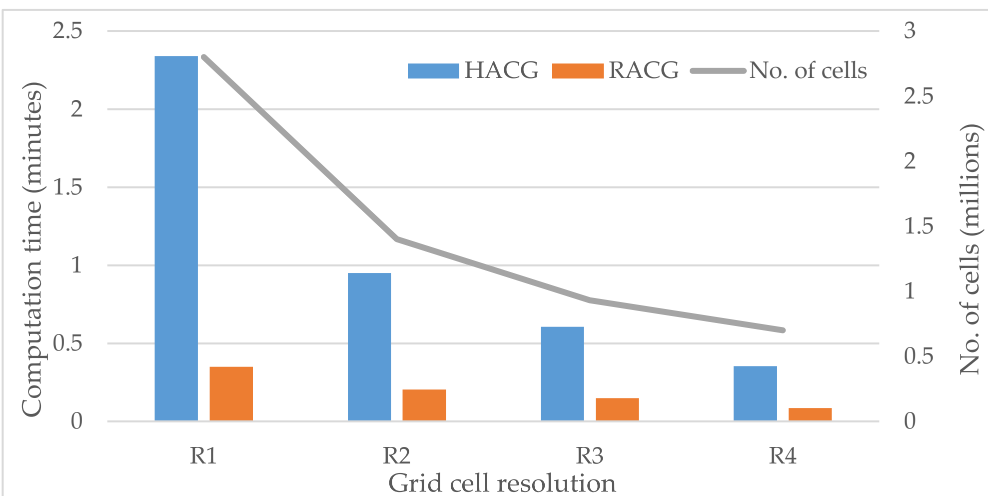


Figure 2.12 The computation efficiency of the raster-based (RACG) and hexagon-based (HACG) adaptive crystal growth Voronoi diagrams for different spatial resolutions. (The left vertical axis is the computation time in minutes, while the right vertical axis is the number of cells for the weighted planes with different resolution.)

2.4. DISCUSSION

The best delineation results for HACG and RACG can be obtained by iterating and calibrating the parameter ω. According to the calibration results of the ω settings, the best observed delimitation results were selected based on how commensurate the population in each service area is with the enrollment capacity of the middle school in the service area and how accessible the middle schools are within their service areas. The delimitation results with observed best ω values are illustrated in Figures 2.11a (RACG) and 11b (HACG). In these figures, the polygons with distinct colors are the service areas of different middle schools. Because the road network, natural barriers, and the population distribution were taken into account through two weighted planes in the adaptive crystal growth Voronoi diagrams, different schools have differently sized service areas; generally speaking, the service area of a middle school is smaller when the road network is sparser, the population density higher, and more natural barriers exist in its surrounding area. Not surprisingly, the shape of some middle schools' service areas is elongated along the transport network. This is because these middle schools are located in sparsely populated areas where the crystal growth is not constrained and the crystal growth speed along transport network cells is faster than in other cells due to the high accessibility. When comparing the delimitation results of the raster-based and hexagon-based adaptive crystal growth Voronoi diagrams, Figure 2.11 highlights the fact that these two service area delimitation methods lead to different results with the same crystal growth rules.

Regarding the ω value, which is the only parameter in the method and represents how restrictive the growth constraint rule is, the maximum value of ω for calibration could be defined by the specific question of how great a percentage of service over its capacity for each facility is allowed and tolerable. For instance, considering the middle school service capacity, we require that the service population should not go over its service capacity by 10%. Otherwise, the school would not be able to provide quality education for students. Based on that, the maximum value of ω could be set as 0.1. Therefore, the value can be calibrated in the range from 0 to 0.1. Further, the process to find the optimal ω value is problem-dependent. It depends on the particular service facility in question, and tolerance of errors of the two measurements for the specific questions of (1) how commensurate the population in each service area is with the enrollment capacity of the corresponding middle school (RMSE or Max. Error); and (2) how accessible each middle school is within its service area (Max. or Mean LLT). If the tolerance of errors for how commensurate the population in each service area is with the enrollment capacity of the corresponding middle school is 10%, and the target is to find the optimal results with the least mean LLT, then the calibration procedure is to find the optimal ω value with smallest mean LLT while the Max. Error is no larger than 10%. If the target is to find the best fit of the service population with the enrollment capacity and the tolerance of travel time in each service area is 25 min, then the calibration procedure is to find the optimal ω value with smallest RMSE while the Max. LLT is no larger than 25 min.

The results prove that HACG generates better delineation results than does RACG in many ways. Considering how commensurate the population in each service area is with the enrollment capacity of the corresponding middle school, Figures 2.6 and 2.7 show that the proportion of the population in different service areas based on the results of the hexagon-based method is more commensurate with the proportion of middle schools' enrollment capacity. This may suggest that the hexagon grid may suffer less from orientation bias and sampling bias and thus can generate service areas that better match the service capacities of the middle schools in the study area. Regarding the accessibility of each middle school within its service area, Figures 2.8–2.10 indicate that the LTT of the service areas delimited by the hexagon-based method is smaller than the LTT

of the service areas delimited by the raster-based method. These results indicate that the hexagon-based adaptive crystal growth Voronoi diagrams produced results superior to those obtained with the raster-based diagrams in addressing both the socioeconomic context and accessibility (travel time based on the road network and natural barriers) when delimiting service areas of public schools. Regarding the spatial resolution of the grid cells, the results indicate that the finer the resolution, the better the delineation results for both RACG and HACG. However, with the same resolution, HACG always performs better than RACG.

Although the proposed method extends the adaptive crystal growth Voronoi diagrams and performs better than the raster-based methods, there are some limitations which should be addressed in future research. First, a considerable amount of inner-city travel is made by public transit (e.g., bus, subway), which is not considered in this research. It should be integrated into the accessibility-weighted plane to better evaluate the accessibility to public facilities. In addition, all areas are considered as walkable area except buildings, rivers, and lakes, and this is not accurate. Routes that are unsuitable for walking were ignored, and pedestrian sidewalks and pedestrian cut-throughs should be considered in future studies. Further, the socioeconomic weighted plane in this study only considered population distribution; weighted planes of other socioeconomic attributes and new crystal growth rules based on multiple socioeconomic weighted planes are also topics for future research. Lastly, this study is a pilot exploration of the method on school service area delineation to justify the advantages of HACG compared with RACG. However, more studies in different locations with different context settings are needed to investigate the merits and shortcomings of HACG further. Moreover, the method should be further examined in a broader set of test problems in order to be generalizable to the service area delineation of other public facilities.

CHAPTER 3: CONTEXT-BASED CRYSTAL GROWTH ACTIVITY SPACE METHOD FOR ENVIRONMENTAL EXPOSURE ASSESSMENT[2]

3.1 INTRODUCTION

In the fields of epidemiology and health geography, understanding environmental exposure is a nontrivial issue involving the investigation of environmental effects on human health. Researchers have examined the relationships among human movement, environmental context, and health outcomes over the past decades. Abundant research has shown that physical activity, tobacco use, obesity, mental health and many other health behaviors or issues are related to environmental exposure (Millstein *et al.* 2009, Epstein *et al.* 2014, Koohsari *et al.* 2015, Shareck *et al.* 2015). For instance, it has been found that the built environment influences physical activity and health (Saelens, Sallis, and Frank 2003, Lee and Moudon 2006)—e.g., obesity is more prevalent in areas that lack physical activity facilities (Giles-Corti *et al.* 2005) or are unfriendly to walking (Ewing *et al.* 2003, Frank *et al.* 2004). One of the fundamental questions in this research area is how to measure environmental exposure.

Despite the existence of many methods, the residential neighborhood is predominantly utilized as the contextual unit for environmental exposure measurement. It is often represented by administration areas, such as census tracts and postal units, because of the availability and easy access to routine administrative data. The readily available spatial delineations of administrative areas and the lack of detailed mobility data are other reasons for the popularity of administrative areas in environmental health research. With the help of advanced geospatial technologies (e.g., geographic information systems [GIS] and global positioning system [GPS]), there has been a

---

[2] Reprint, with permission, from Wang, J., Kwan, M.-P., and Chai, Y., 2018. An Innovative Context-Based Crystal-Growth Activity Space Method for Environmental Exposure Assessment: A Study Using GIS and GPS Trajectory Data Collected in Chicago. International Journal of Environmental Research and Public Health, 15 (4), 703–726.

methodological shift in health research, moving from using fixed administrative areas as contextual units to ego-centered definitions (Miller 2007, Lee *et al.* 2008, Chaix *et al.* 2009). An ego-centered neighborhood is usually represented by a buffer area centered on an individual's home with a given threshold of specific distance or travel time (Perchoux *et al.* 2013), which may reflect more accurately the exposure area rather than administrative units. Due to the ongoing debate as to the best way to define geographic context (Weber and Kwan 2003, Inagami *et al.* 2007, Saarloos *et al.* 2009, Kwan 2012a) and the availability of activity diary and GPS tracking data, many researchers have now adopted the idea that the residential neighborhood can only partially capture people's exposure to environmental context, and daily activities that take place at other locations also contribute to their environmental exposures (Chaix *et al.* 2009, Rainham *et al.* 2010, Houston 2014). The shift from a static measuring approach to a dynamic one has inspired researchers to explore and develop exposure assessment methods using individual GPS tracking data (Duncan *et al.* 2009, Maddison and Ni Mhurchu 2009, Chaix, Méline, Duncan, Jardinier, *et al.* 2013).

Although there are many ways to measure environmental exposure, activity space based on GPS tracking data (movement data) appears to be a promising way to assess the environment utilized by individuals, and to which they are exposed (Krause 2012, Shen and Chai 2013). Activity space is defined as "the local areas within which people move or travel in the course of their daily activities" (Albert and Gesler 2003). Because activity space indicates where and how people have contact with their social and physical environments (Golledge 1997), it can be used as a measure of "people's degree of mobility" (Gesler and Meade 1988). The activity space of an individual can thus be used to explore the interaction between human activity and environmental context (Sharp *et al.* 2015, Tamura *et al.* 2017). Standard deviational ellipses, GPS trajectory buffers, minimum convex polygons, and kernel density surfaces have been widely used to represent human activity space (Cummins 2007, Perchoux *et al.* 2013, Sharp *et al.* 2015). Notwithstanding the improvements in the theory and methodology to assess environmental exposure and in the investigation of the contextual effects on health outcomes with activity space, substantial challenges remain.

Even with advanced activity space methods to assess individual environmental exposure, inconsistent findings of the environmental effects on health behaviors/outcomes have been observed in recent studies (Diez Roux 2001, Oakes *et al.* 2007, Adams and Kapan 2009). This suggests that the reliability of existing studies may be affected by the misspecification of the geographic context (Spielman and Yoo 2009), which was recently articulated as the uncertain geographic context problem (UGCoP) by Kwan (2012a). The UGCoP refers to the problem that findings of the effects of area-based environmental variables (e.g., land-use mix) on health outcomes or behavior (e.g., physical activity) can be affected by how contextual units are geographically delineated. The problem "arises because of the spatial uncertainty in the actual areas that exert contextual influences on the individuals being studied and the temporal uncertainty in the timing and duration in which individuals experienced these contextual influences" (Kwan 2012b). Existing activity space methods have limitations that may compromise their ability to mitigate the UGCoP both spatially and temporally. From the perspective of spatial uncertainty, existing methods ignore the accessibility of different locations in the study area and thus may include locations that may not be accessible to people. Moreover, arbitrary cut-off distances are often used for delineating activity space. Temporally, the duration of environmental exposure is treated merely as the multiplier of exposure while individuals' interactions with space during particular periods of time (the more time spent at the location, the more familiar with the sounding area) is not considered.

This paper proposes the context-based crystal-growth (CCG) activity space as an innovative method for generating individual activity space based on both GPS tracking and environmental context. To mitigate the UGCoP, portable GPS devices are utilized to trace human movement accurately, and advanced GIS methods are used to relate these data to high-resolution data of relevant environmental contexts (Almanza *et al.* 2012, Kwan 2012a). The integration of GPS and GIS provides a powerful means for examining the relationships between environmental contexts and health outcomes (Wiehe, Hoch, *et al.* 2008, Maddison and Ni Mhurchu 2009). In contrast to other existing methods, and in order to address spatial uncertainty, activity space is generated considering not only people's actual daily activity patterns based on GPS tracks but also the environmental contexts that either constrain or encourage people's daily activity. Instead of using arbitrary cut-off distance, activity space is delineated based on the features of individual movement patterns. To mitigate temporal uncertainty, the duration of the environmental context in which individuals experienced and with which individuals interact are taken into account by abstracting the core areas of their daily activities. The size of activity space is based on the accumulated time an individual spent at the location (the more time a person spent there, the larger the activity space). To the best of our knowledge, this is the first study to introduce the context-based crystal-growth method and consider both people's daily activity patterns and environmental context in activity space and environmental health research. The results indicate that the proposed new method generates more reasonable activity space and more accurate exposure assessment when compared to other existing methods. It can help mitigate the UGCoP spatially and temporally in environmental exposure measures. The accurate assessment of environmental exposures sheds light on the investigation of environmental effects in the field of epidemiology and health geography. The method is a new tool for activity space delineation and can be used for exploring the relationships among human movement patterns, environmental context, and health outcomes.

3.2 METHODOLOGY

This research proposes an innovative method for delineating individual activity space based on GPS tracking data, accessibility-related contextual data, and a crystal-growth algorithm. Due to the capability of incorporating GPS tracking and contextual data, as well as the flexibility to adaptively adjust the activity space based on the context of accessibility, this method is suitable for generating activity space while mitigating the UGCoP. In this method, accessibility-related contexts are incorporated into weighted planes, in which space is tiled into fine regular-grid cells (e.g., hexagonal cells). The weighted planes use hexagon grids to achieve higher accuracy in representing the spatial features of the land surface and minimize orientation bias and sampling bias from edge effects. The method is also capable of handling different transportation modes (e.g., walking, driving, taking the bus or train) while generating the activity space. Two accessibility-weighted planes are generated for public transport users and private transport users, respectively, considering the different effects of context for various groups of residents. Based on the weighted planes, space is delineated by the growth of cells from one or more seed points to neighbor cells through a sequence of growth cycles. The crystal-growth method is suitable for this study because the growth speed of each cell can be dynamically adjusted according to the accessibility-weighted planes, and the growth extent can be feasibly defined based on travel time. Figure 3.1 illustrates the workflow of the proposed context-based crystal-growth method. As people's frequently visited locations are essential for understanding their daily activity, these places are considered as the core areas of their activity space, as identified by the kernel density analysis of an individual's 7-day

GPS trajectories. Thus, the activity space is grown from the core areas, and they will grow to their neighbor cells cycle by cycle. The crystal-growth process is either constrained or encouraged based on the accessibility-weighted planes. The merging of the crystal-growth space of all core areas generates the activity space.

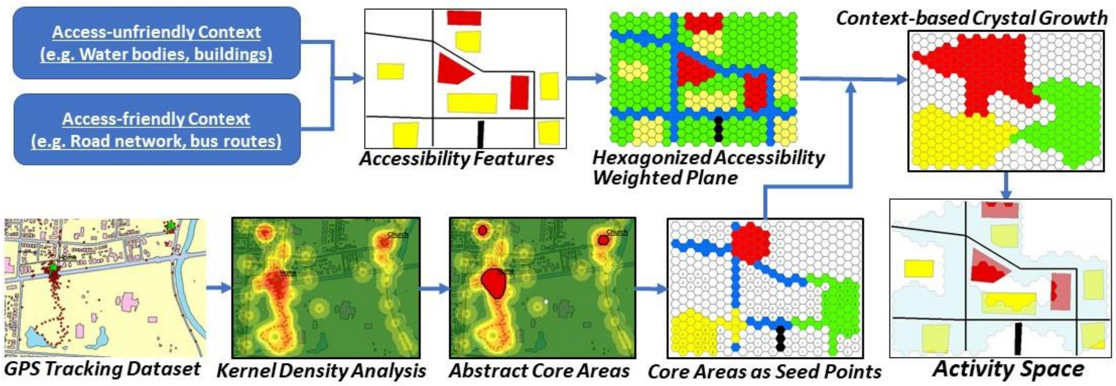


Figure 3.1 Workflow of the context-based crystal-growth activity space method.

3.2.1 GPS Tracking and Context Data

The individual GPS tracking dataset used in this research was collected as part of a larger study that examines the relationship among the exposure to environmental stressors, neighborhood quality, and individual health in Chicago. The larger study seeks to understand how the neighborhoods people live in and visit in their daily life affect their health and wellbeing. It focuses on the noise and air quality that people are exposed to in their daily life—not just at their residence, but also while they undertake their daily activities at other locations (e.g., travel to work, shopping, or running errands). The data were collected from October to December 2017 in the Chicago metropolitan area using surveys, GPS-equipped mobile phones, and portable noise and air pollutant sensors. The GPS tracking dataset is not recorded in even temporal duration but somewhat random over time depending on participants' movement for prolonging battery life. To be specific, if a subject moves more than 1 m from the previous record within 3 s, a tracking point will be recorded by the tracking device. If the subject does not move more than 1 m from the last record in 3 s, a new tracking point is still recorded. In the GPS tracking dataset, each subject was tracked with GPS-equipped mobile phones for seven days. Because most people have highly routinized daily activities (Kestens *et al.* 2018), 7-day continuous activity tracking, which covers both weekdays and weekends, can capture most of the participants' weekly routine activities and is typically used for activity space research (Zenk *et al.* 2011, Shen *et al.* 2013, 2015, Ta *et al.* 2015, Kestens *et al.* 2016, 2018, Tana *et al.* 2016). Consistent with previous studies, GPS tracks for 7 consecutive days were used to generate individual activity space in this study.

To compare the activity space generated by different methods, as some of them are sensitive to participants' movement patterns, four representative participants are selected from the dataset, as their movement trajectories have very different patterns (the spatial arrangement of the GPS points). The movement trajectories of Person A show a compact clustered pattern, those of Person B exhibit a modest clustered pattern, those of Person C display a one-directional pattern, and those of Person D present a multi-directional pattern. Note that due to the Institutional Review Board's (IRB) requirements for protecting data confidentiality and participants' privacy, we can describe the patterns but cannot include these maps in this paper.

Further, the activity spaces are used to assess environmental exposure with the whole dataset. The exposure to physical-activity-friendly contexts is used as a proxy for the comparison of activity space methods. Physical activity has been intensively studied in environmental health research (Rodriguez *et al.* 2005, Wheeler *et al.* 2010, Rodríguez *et al.* 2012) because it is highly related to many chronic diseases, such as type-II diabetes, obesity and cardiovascular diseases (Ewing *et al.* 2003, Lahti *et al.* 2014, Arem *et al.* 2015, Moore *et al.* 2016, O'Donovan *et al.* 2017). Although the results are inconsistent, most previous studies have observed a positive association between physical-activity-friendly environments and the level of physical activity (Handy *et al.*

2002, Mota *et al.* 2005, Berke *et al.* 2007, Maroko *et al.* 2009, Santos *et al.* 2009, Zenk *et al.* 2011, Koohsari *et al.* 2015, Roberts 2017); the effectiveness of activity space for environmental exposure assessment can be evaluated by examining whether the association between exposure to the context and physical activity is captured. Physical-activity-friendly contexts include green spaces, blue spaces and other leisure facilities such as urban parks, playgrounds, swimming pools, and sports centers. The level of physical activity, which was reported by participants in the questionnaire before the GPS tracking, is measured by the number of days over a typical week in which a participant is physically active for a total of at least 30 min. There are 31 participants whose GPS tracking data in the dataset are used for such evaluation. The sociodemographic characteristics of these participants are shown in Table 3.1. The physical-activity-friendly contexts are just an example used to illustrate the usefulness of the proposed activity space method for environmental exposure assessment, which can also be applied to examine the effects of other environmental influences on other kinds of health behaviors or outcomes. For example, to explore the environmental influences on people's body weight, we can include the availability of different types of food outlets or shops (e.g., fast-food outlets that sell unhealthy foods and supermarkets that provide more selections of healthy foods. To examine people's exposure to air pollution, we can include traffic-related sources (e.g., highways) and various point sources (e.g., industrial plants; oil refineries) as environmental influences.

Table 3.1 Sociodemographic characteristics of the participants.

Sociodemographic Variables		**Percentage**
Gender	**Male**	**61.3%**
	Female	**38.7%**
Age	**18–30**	**17.2%**
	31–40	**17.2%**
	41–50	**24.1%**
	51–65	**41.4%**
Race	**White**	**9.7%**
	African American	**41.9%**
	Latino/Hispanic	**41.9%**
	Other	**6.5%**
Education	**Elementary School**	**6.5%**
	High School	**58.1%**
	College/University	**32.3%**
	Graduate School	**3.2%**
Marital Status	**Married**	**17.9%**
	Divorced	**14.3%**
	Single	**67.9%**
Annual Income	**Less than $10,000**	**58.1%**
	$10,000–$24,999	**19.4%**
	$25,000–$50,000	**9.7%**
	$50,000–$99,000	**9.7%**
	$100,000 or more	**3.2%**

To generate a consistent and consecutive time series for the GPS trajectories, data cleaning and interpolation of GPS points was conducted so that there is one GPS tracking point each second for every participant for the 7-day tracking period. A Python program is developed to test every consecutive pair of GPS records and calculate the time difference. If the time difference between two consecutive GPS records is N seconds, which is longer than 1 s and shorter than 1800 s (half hour), linear interpolation in both spatial and temporal dimension is performed between the GPS records so that N-1 more records are inserted. After the interpolation process, each participant has one GPS tracking point for every second for the entire tracking period.

The environmental context data for this study were derived from a comprehensive digital geographic database of Chicago from the Chicago Data Portal as well as the volunteered geographic information website of OpenStreetMap. This includes the geographic location and footprint of buildings, water bodies, woods, restricted area (e.g., airport), and ground railways, which can be considered as barrier factors for accessibility; while the various levels of road networks, walkable areas, public transport routes are considered as the access friendly context that encourages accessibility. These contextual data are used to generate the two context-based hexagonal accessibility-weighted planes.

3.2.2 Core Areas of Daily Activities

People's frequently-visited locations are crucial for understanding their daily activity and the delineation of their activity space. Therefore, these locations (core areas) are first abstracted from the GPS trajectories using kernel density analysis. The results of the kernel density analysis of the GPS trajectories are not distributed normally and skewed to the low-density values, so the geometric interval classification is a suitable method for classifying the density values by minimizing the sum of squares of the number of elements in each class. The algorithm creates geometric intervals to ensure that each class range has approximately the same number of values while keeping the change between intervals consistent (ESRI 2017). The core areas are abstracted from the results of the kernel density analysis as the collection of cells with density value larger than the 3/4 cut point of the geometric interval classification of all non-zero values. These core areas will be used as the seed points in the following crystal-growth delineation. Further, the number of the growth cycle of each core area is weighted based on the accumulated time the individual spent at these locations. The more time a person spent there, the more cycles the activity space grows. The crystal-growth cycle of each core area is calculated based on the normalized sojourn time (NST) at each core area. For instance, one individual spent 600 min daily at home on average, so the normalized sojourn time at home (NST_h) is 600. The growth extent (GE) is dependent on the NST. We assume the growth extent from home location is GE_h, so the GE_i of other core areas can be calculated based on the following formula:

$$a = \sqrt[GE_h]{NST_h} \tag{3.1}$$

$$GE_i = \log_a NST_i \tag{3.2}$$

where GE_i is the growth extent of seed point i in minutes, NST_i is the normalized sojourn time at core area i, NST_h is the normalized sojourn time at home, GE_h is the growth extent from the home location.

As the only parameter for defining the growth extent of all the core areas, the GE_h can be fixed universally for all subjects, and it can also be determined based on personal mobility. For the purpose of illustration, this study uses 10-mins' travel as the universal value of GE_h. Although the GE_h could be different for people with different mobility, 10-mins' travel distance is a reasonable assumption that people are familiar with the environment and activity opportunities of the areas around the home, and it is highly possible that people choose to undertake the daily activity and are exposed to the context in this area.

3.2.3 Context-Based Hexagonal Accessibility-Weighted Planes

The accessibility-weighted plane, as a representation of the accessibility-friendliness of the environmental context, considers the effect of many kinds of environmental contexts (e.g., rivers as barriers, roads as thoroughfares) on the accessibility of the study area. The context of buildings, water bodies, woods, restricted areas, ground railways, road networks, public transport routes (metro and bus) and walkable areas are critical factors for the accessibility-weighted plane. Further, considering the different effects of environmental contexts for various groups of people (e.g., the expressway is considered as a thoroughfare for car users, while it is a barrier for public transport users), two context-based hexagonal accessibility-weighted planes are generated respectively for private transport users and public transport users.

For private transport users, on the one hand, among all these contexts, buildings, water bodies, woods, restricted areas (e.g., airport, private land) and ground railways are considered as barrier factors that are normally hard to trespass by people. On the other hand, various levels of road networks and walkable areas are accessibility-friendly context since they increase the general approachability of various sites. Road networks are further classified into expressways, primary roads, secondary roads and tertiary roads with various travel speeds. On the accessibility-weighted plane, as illustrated in Table 3.2, seed point cells are assigned a value of 100. The barrier cells (e.g., restricted areas and water bodies) are designated a value from 10 to 14 with a growth speed of 0. The transport network cells are assigned a value of 21 to 24, and the growth speed varies with their average travel speed. For pedestrian trails and walkable areas, cell values of 30 and 31 are assigned, respectively, and the growth speed is 1 cell per cycle (3 miles/h).

Table 3.2 Cell attributes of the hexagonal accessibility-weighted plane for private transport users.

Context Type	**Cell Value**	**Average Moving Velocity**	**Growth Speed (Cells/Cycle)**
Out of Research Area	0	-	0
Restricted Area	10	-	0
Water Bodies	11	-	0
Woods	12	-	0
Buildings	13	-	0
Railway	14	-	0
Expressway	21	About 35 miles/h	12
Primary Road	22	About 25 miles/h	8
Secondary Road	23	About 15 miles/h	5
Tertiary Road	24	About 9 miles/h	3
Pedestrian Trail	30	About 3 miles/h	1
Walkable Area	31	About 3 miles/h	1
Seed Points	100	About 3 miles/h	1

In contrast to private transport users, for public transport users, expressways, buildings, water bodies, woods, restricted areas, and ground railways are considered as barriers, while metro routes and bus routes are considered as thoroughfares with higher accessibility. On the weighted plane for public transport users, all kinds of local road (including primary roads, secondary roads, tertiary roads, and pedestrian trails) are considered as walkable area only, since they do not have cars to drive on them. As listed in Table 3.3, local roads and walkable areas are assigned a value of 30 and 31, with a growth speed of 1 cell per cycle. Bus routes have a growth speed of 3 cells per cycle with a cell value of 23, while metro routes have a growth speed of 6 cells per cycle with a cell value of 21. Metro stations are also marked in the weighted plane (cell value: 22; growth speed: 1) since citizens can only get on or get off the metro system at stations. On the contrary, bus stations are not considered, since buses stop frequently and the distance between bus stations are only about 100 to 200 m in Chicago.

Table 3.3 Cell attributes of the hexagonal accessibility-weighted plane for public transport users.

Context Type	Cell Value	Average Moving Velocity	Growth Speed (Cells/Cycle)
Out of Research Area	0	-	0
Restricted Area	10	-	0
Water Bodies	11	-	0
Woods	12	-	0
Buildings	13	-	0
Railway	14	-	0
Expressway	15	-	0
Metro Routes	21	About 18 miles/h	6
Metro Stations	22	About 3 miles/h	1
Bus Routes	23	About 9 miles/h	3
Local Roads	30	About 3 miles/h	1
Walkable Area	31	About 3 miles/h	1
Seed Points	100	About 3 miles/h	1

The geographical location and footprint of the environmental contexts are utilized to generate the hexagon-grid-based accessibility-weighted plane. The hexagonal grid, different from the raster grid, tiles the land surface with regularly sized hexagonal cells. They are the most compact regular polygons that can fill the land surface (Birch *et al.* 2007). Hexagonal cells are closer in shape to circles (Feick and Robertson 2015, Zook 2015), and they have only one kind of neighbor cells that share the same edge. Further, the distance to the centroid of a cell from the six neighboring cells is the same. The hexagon grid can achieve high accuracy in representing the spatial features of land surface from the perspective of spatial analysis (Zook 2015), and it reduces the complexity in defining neighbor cells when compared to the raster grid. Thus, it is suitable for the crystal-growth algorithms used in this study. Figure 3.2 illustrates the environmental contexts of a neighborhood in Chicago, and the generated hexagonal accessibility-weighted plane for private transport users (see Figure 3.3) and public transport users (see Figure 3.4). The hexagonal cells in the weighted planes have a fine resolution of 10 × 10 m to ensure the accuracy of the calculation and were generated using ArcMap. The context-based hexagonal accessibility-weighted planes of Chicago are shown in Figure 3.5; there are about 6 million hexagonal cells for each weighted plane that covers the city.

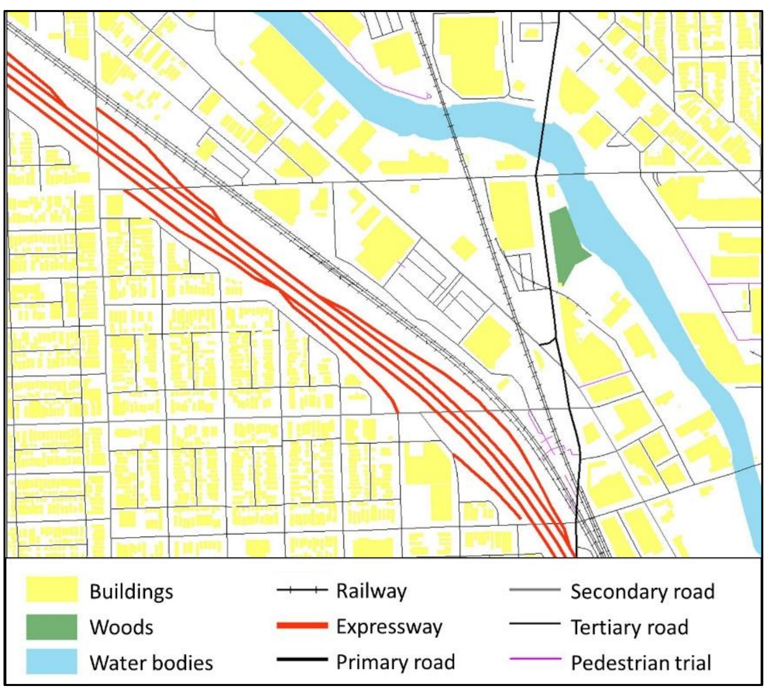


Figure 3.2 Accessibility-related environmental contexts.

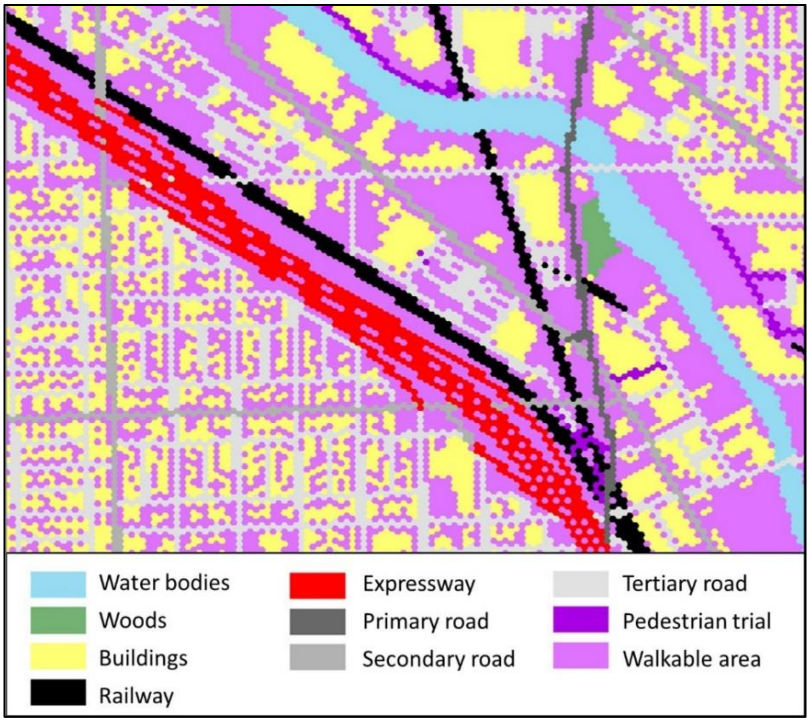


Figure 3.3 Hexagonal accessibility-weighted plane for private transport users.

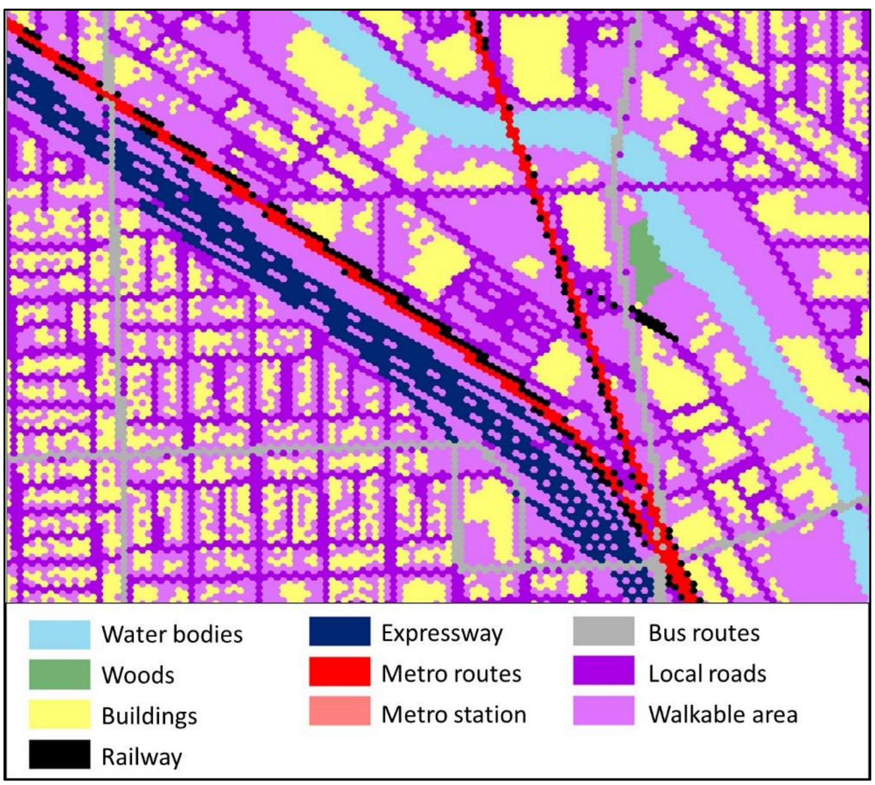


Figure 3.4 Hexagonal accessibility-weighted plane for public transport users.

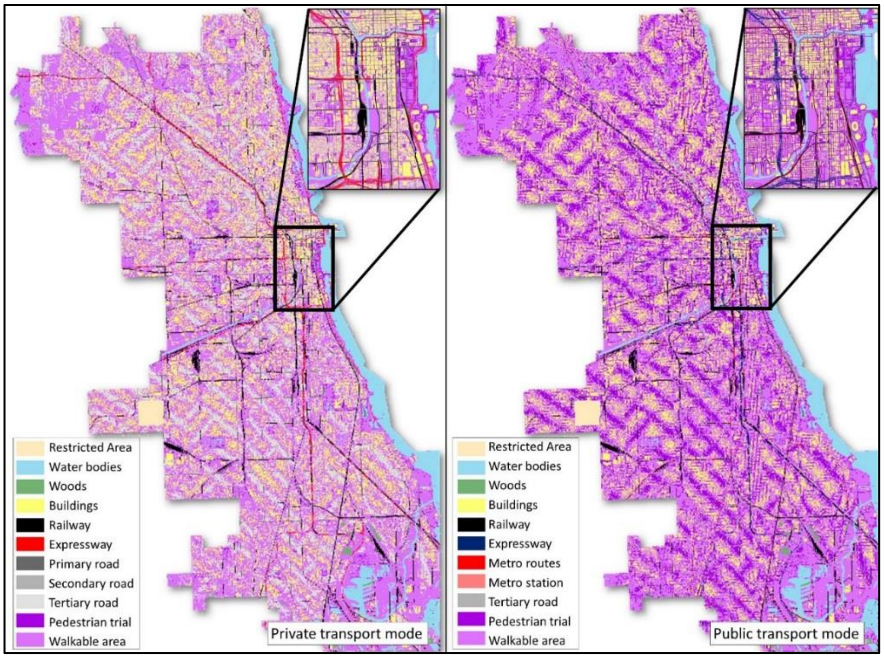


Figure 3.5 The context-based hexagonal accessibility-weighted plane of Chicago.

3.2.4 Hexagon-Grid Crystal-Growth Activity Space

The crystal-growth method is a tool for space partitioning and was used for Voronoi diagrams (Zook 2015), spatial delineation (Wang *et al.* 2014), and spatial optimization (Lin and Huang 1998). The method is suitable for this study because the growth speed of each cell could be adjusted based on the context of accessibility. Since the context of accessibility is one of the critical factors for accurately generating activity space, which is rarely considered in previous research, this study utilizes the crystal-growth algorithm based on the hexagon-grid accessibility-weighted planes.

Figure 3.6 illustrates this crystal-growth algorithm. In the method, the activity space is grown from all seed point cells (the location of the core areas), the service area of each seed point cell will grow to their neighbor cells cycle by cycle. Because the weighted planes simulate the road network and other physical barriers (Figure 3.6a), the growth speed of each location can be adjusted in real time based on the attributes of the cells. For instance, roads speed up growth, while rivers or lakes block the growth. For illustration, the road network is rendered as blue cells, while black cells represent barriers in the figure. As illustrated, crystal-growth starts from the seed point cells to their six immediate neighbor cells, and the growth continues cycle by cycle (Figure 3.6a,b illustrates one crystal-growth cycle). In contrast to other cells, the growth speed of transport network cells is faster due to the higher accessibility they enable (Figure 3.6b,c). Growth is constrained by natural barrier cells because it is usually challenging to travel through barriers such as rivers and lakes (Figure 3.6c,d). The growth of each seed point will stop when its maximum growth extent reached. The merging of the crystal-growth area of all core areas generates the final activity space.

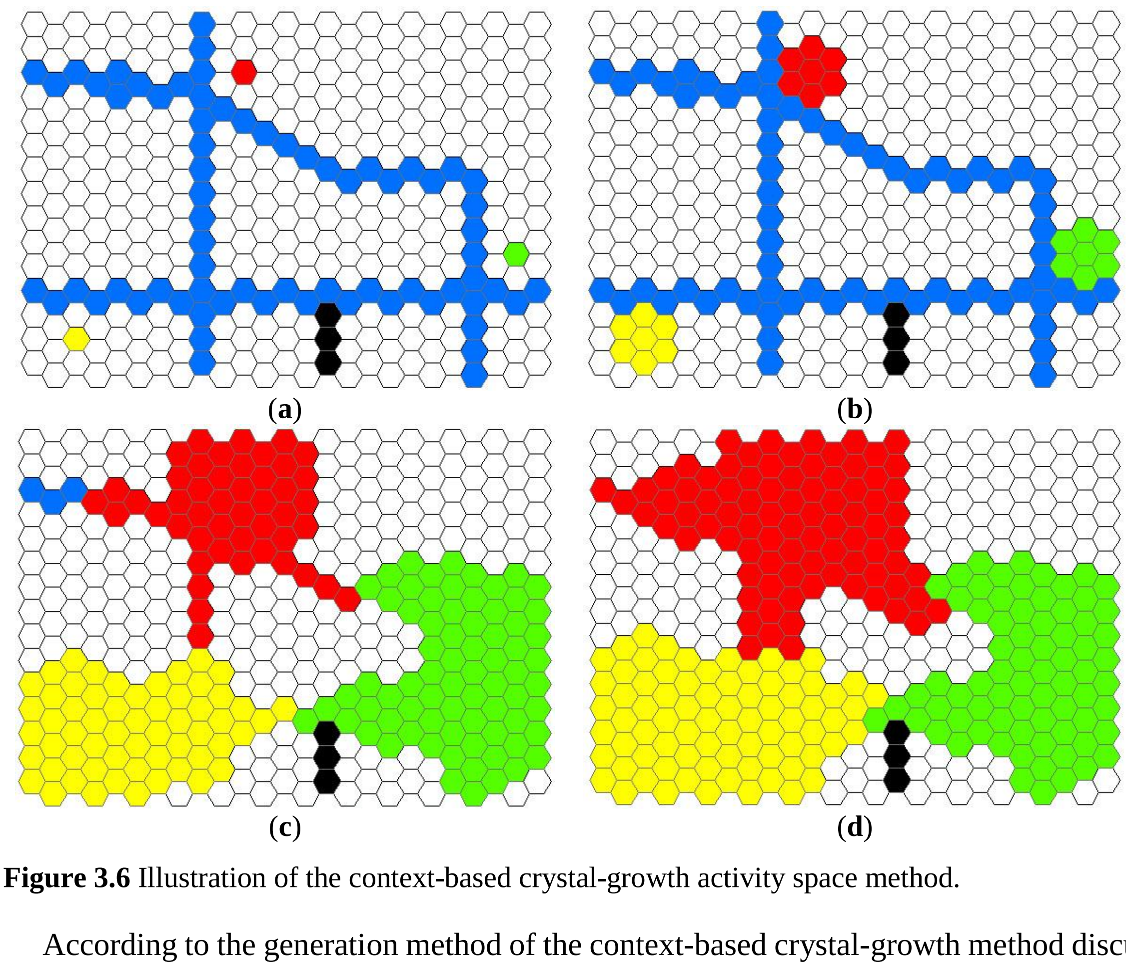


Figure 3.6 Illustration of the context-based crystal-growth activity space method.

According to the generation method of the context-based crystal-growth method discussed above, activity space is generated according to the following growth rules. (1) Based on participants' travel mode, choose the context-based hexagonal accessibility-weighted plane accordingly for either private transport users or public transport users. (2) The crystal-growth starts from all the seed points (core area cells) separately. Seed point cells are the hexagon cells that intersect with the locations of the centroids of core areas. (3) For each crystal-growth cycle, if a neighbor cell is a barrier cell or a cell that has already been labeled as a grown area, it will be skipped. Otherwise, the neighbor cell will be marked as the grown area. (4) The crystal-growth speed of each cell is determined by the cell value, which was defined on the accessibility-weighted planes. (5) The crystal-growth will finish if the growth of all the seed cells reached their maximum growth extent. (6) The final activity space is generated by merging all the crystal-growth areas.

3.2.5 Other Existing Activity Space Methods

To compare the proposed context-based crystal-growth (CCG) activity space method with other existing methods, four commonly used delineations of activity space are implemented with the same GPS dataset, including GPS trajectory buffers (GTB), standard deviational ellipses (SDE), kernel density surfaces (KDS) and minimum convex polygons (MCP). GTB was created for each selected participant by covering the subject's GPS trajectories with a 200-m buffer area, which covers all the locations that a participant visited or passed by during the study period. SDE is a commonly used method for delineating individual activity space. The SDE captures the geographic distribution and directional trend of all activity locations. The ellipse was obtained based on one or two standard deviations of the distances between each point and the transformed mean center along the rotated major and minor axes of the point set. Since past studies have used either one standard deviation SDE (SDE1) or two standard deviations SDE (SDE2), we derived both in this study for comparative purposes. KDS is a density surface derived from the activity locations and an associated weight using a kernel function and a predetermined search radius. In this study, the KDS was generated based on the duration spent at each GPS point as the weight on a raster layer at the spatial resolution of 10 × 10 m and search radius of 1000 m. MCP for a subject is the smallest convex polygon that contains all of the person's GPS tracking points. It represents the smallest area that includes all activity destinations of a participant.

3.3. RESULTS

3.3.1 Context-Based Crystal-Growth Activity Space

The CCG activity spaces delineated based on the four selected participants are illustrated in Figure 3.7. For person A, whose daily activity is highly concentrated around the home location, it shows a compact clustered pattern. Based on the collected demographic characteristics, this person is a retired female in her 60s, who spends most of her time at home and only visits the public library regularly. Only two core areas are detected, and they are close to each other. Thus, the CCG activity space is a grown area centered at these two locations within the travel distance of 10 min from home (one of the core areas) and the corresponding travel distance (calculated based on the normalized sojourn time) from the other core area (library). Because the growth is based on the accessibility-weighted plane, the activity space protrudes along major roads due to their higher accessibility than other contexts. There are several hollow areas in the activity space, which are barrier contexts, such as water bodies and private houses, that could not be easily trespassed. In contrast to the compact clustered pattern, person B presents a modest clustered pattern. Although the movement is also clustered around the home, this person has more activity locations and travels much further than person A to undertake daily activities. According to the participant's profile, this subject is an unemployed male in his 40s. He visits his mother's home frequently in the west of the city about half-hours' drive from home. He also visits the downtown area to visit doctors and friends. Not surprisingly, three core areas are identified, and his CCG activity space is composed of three separate grown areas centered at home, mother's house and downtown. The largest sub-area is the one centered at home (in the middle of the map) based on the fact that he spends the largest amount of time at home. He also spends much time at his mother's house, and the sub-area should be larger than the current form, which is truncated due to edge effects. His mother's home is close to the boundary of the study area, and thus only part of the sub-area is captured due to the lack of contextual information outside the border of the study area. Since the time he spends in the downtown area is much less when compared to the time spends in the other two core areas, this sub-area is much smaller. With a one-directional pattern, the CCG activity space of person C is exhibited in Figure 3.7. This person is a middle-aged male. He does grocery shopping and other personal activities around the home neighborhood. That is probably the reason why many core areas are identified around his residential neighborhood. What's more, with his close relatives living in the southeastern part of the city, he needs to drive there and visit them regularly, and that is why we find another core area there. Thus, his activity space has a significant part centered at home and another small portion in the southeastern part of the city. Finally, person D is a female adult, who has a full-time job at downtown. With considerable mobility and travel around the city for daily activities, this participant's movement shows a multi-directional pattern. As shown in the figure, the two largest portions of her activity space are the ones around home and workplace, since a significant amount of time is spent at these two locations. Other small parts of her activity space are scattered around the city for different daily activities. As a married woman, she needs to take care of the family, which includes grocery shopping and other household-related activities. It is noted that even though she traveled to the further north of the city, no activity space is identified, because she only spends limited time there, and it is not recognized as a core area. It is reasonable to exclude this kind of location from the activity space if the subject only visits the place with limited time on a nonregular basis.

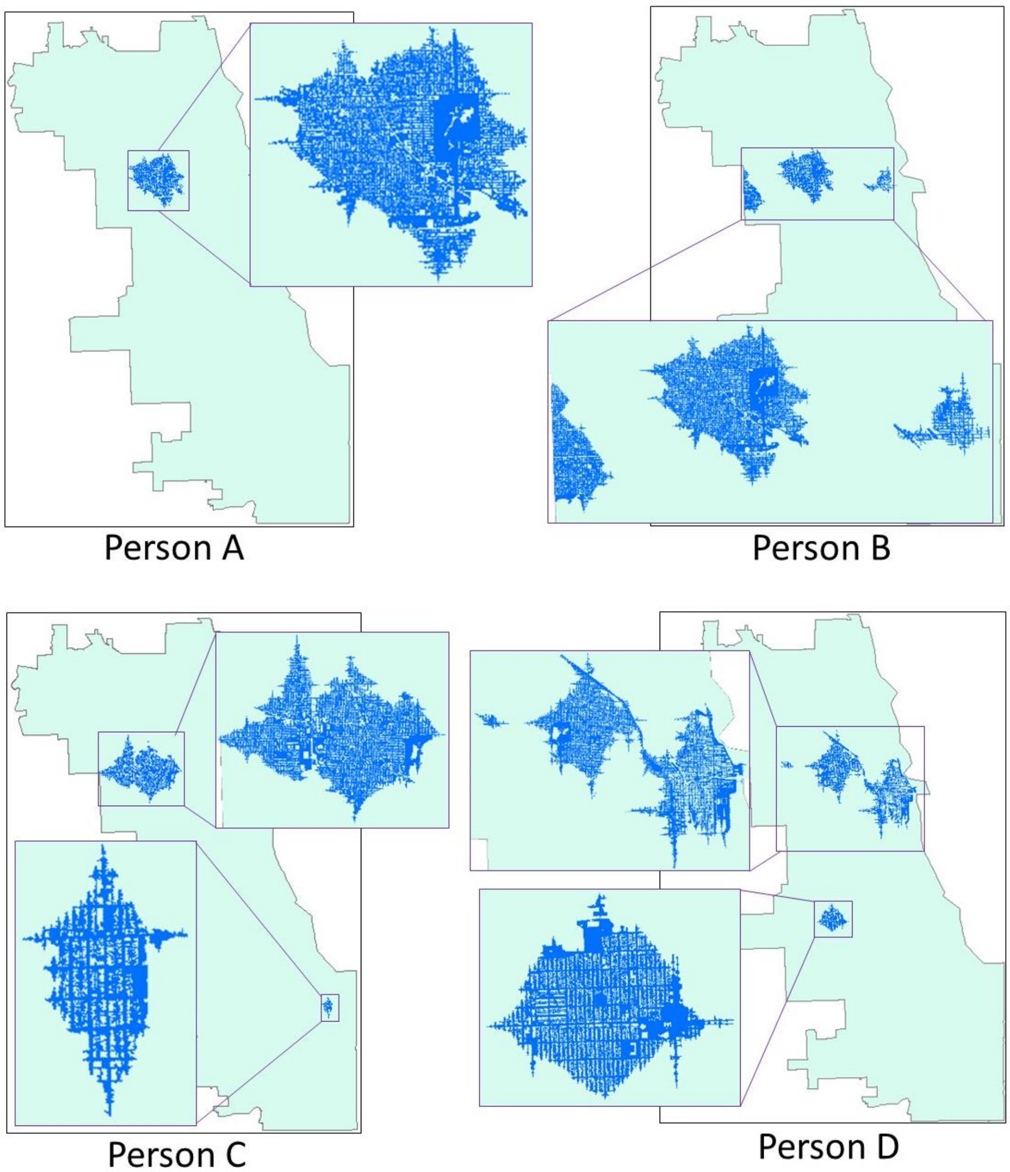


Figure 3.7 The context-based crystal-growth activity spaces of the four representative participants.

To compare and illustrate the difference of CCG based on accessibility-weighted planes for private transport users and public transport users, respectively, as shown in Figure 3.8, both activity spaces are generated with the same core areas at the same locations. For private transport users, the activity space is grown and enlarged along primary and secondary roads. On the contrary, the public transport users relied heavily on the bus and metro systems, so the activity space extends along the bus and metro routes instead of the roads. It can be seen from the figure that the CCG activity space of public transport users is much smaller than that of private transport users.

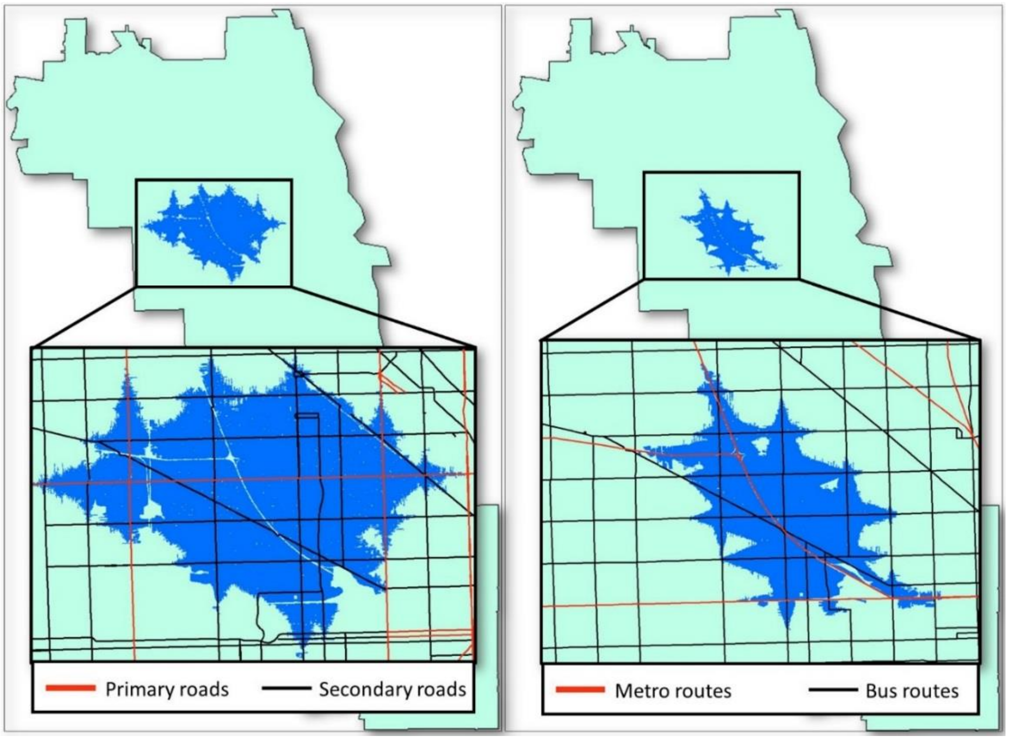


Figure 3.8 Crystal-growth activity spaces based on accessibility-weighted plane for private transport users (**left**) and public transport users (**right**).

3.3.2 Comparing the CCG Activity Spaces with Other Activity Spaces

3.3.2.1 Comparing the activity spaces

The commonly used methods for activity space delineations are implemented with the four representative participants' GPS tracking trajectory to compare with the proposed CCG method. The comparison is based on the geometric characteristics of the activity spaces, matching them with subjects' daily activity, and visual interpretation. The activity spaces of the four participants are illustrated in Figure 3.9, and the comparison results are listed in Table 3.4. Note that the resulting maps of GTB, due to the risk of re-identification for the subjects, are not included in

Figure 3.9 in order to protect their privacy. The activity spaces of person A are similar among the different methods because of the simplicity of the daily movement. Since person A has low mobility, and the GPS trajectories have a compact clustered pattern, all five methods generate activity spaces that are nearly circular, while focusing on the home location of the subject. The area of activity spaces generated by CCG (8.21 km^2) and KDS (6.58 km^2) are significantly larger than the area of activity spaces generated by the other two methods, while they cover almost all of the total GPS tracking points. For more complicated movement patterns (persons B, C, and D), the CCG is capable of generating multiple areas to portray individual activity space, whereas all the other methods only create a single continuous region. The area of the activity space of person B is larger than the one of person A. KDS (66.66 km^2), and MCP (65.52 km^2) have the largest area, while CCG (11.74 km^2) has the smallest area. CCG activity spaces have the highest coverage of the total GPS tracking points except for the GTB, KDS, and MCP. For person C, the GTB, KDS, and MCP include not only the area for daily activity, but also the places along the travel trajectories between activity locations, leading to their larger areas. The directional pattern of the GPS trajectories makes the SDE highly compressed. Although the SDE does not include the places along the travel trajectory, it ignores the activity location in the southeastern part of the city and covers a lot of unrelated sites in the northwestern part. Still, the CCG activity spaces have the smallest size among all the activity spaces (18.76 km^2) for person D. The dispersed GPS trajectories around the city make the area of the activity spaces generated with KDS, MCP, and SDE extremely huge, which covers many irrelevant city spaces. On average, the CCG has the smallest area (12.65 km^2) among all the methods and the highest coverage (94.04%) of total GPS tracking points except for GTB, KDS, and MCP. Because of the nature of the techniques themselves, GTB, KDS, and MCP always cover all the GPS tracking points. On the contrary, CCG, and SDE include only the prominent parts of the points since they are more focused on the characteristics of the movement patterns instead of every single GPS point.

Table 3.4 Comparing the representative participants' activity spaces generated by different methods.

		CCG	GTB	KDS	MCP	SDE1	SDE2
Person A	Area (km^2)	8.21	0.91	6.58	0.52	0.0373	0.15
	%GTPC	100%	100%	100%	100%	93.28%	93.45%
Person B	Area (km^2)	11.74	31.14	66.66	65.52	12.704	50.82
	%GTPC	87.97%	100%	100%	100%	76.47%	85.71%
Person C	Area (km^2)	11.89	19.14	81.20	105.10	7.71	30.83
	%GTPC	96.60%	100%	100%	100%	82.04%	93.72%
Person D	Area (km^2)	18.76	28.92	101.28	109.16	39.12	156.48
	%GTPC	91.69%	100%	100%	100%	75.69%	90.83%
Average	Area (km^2)	12.65	20.03	63.93	70.08	14.89	59.57
	%GTPC	94.04%	100%	100%	100%	81.87%	90.93%

Note: the area of KDS is measured as the total area with positive density, which is consistent with Schonfelder and Axhausen (2003); %GTPC: percent of total GPS tracking points covered.

For a more comprehensive comparative analysis, the activity spaces of the 31 participants in the dataset are also generated by CCG and the other five methods. Figure 3.10 illustrates the differences in the size of the activity spaces for private transport and public transport users. Considering the median size of the activity spaces, private transport users have larger activity spaces than those of public transport users for all the six methods; while the average size of the activity spaces of private transport users is still larger when compared to that of public transport users for all the methods except MCP. However, as shown in the boxplot in Figure 3.10, the difference is only significant for CCG.

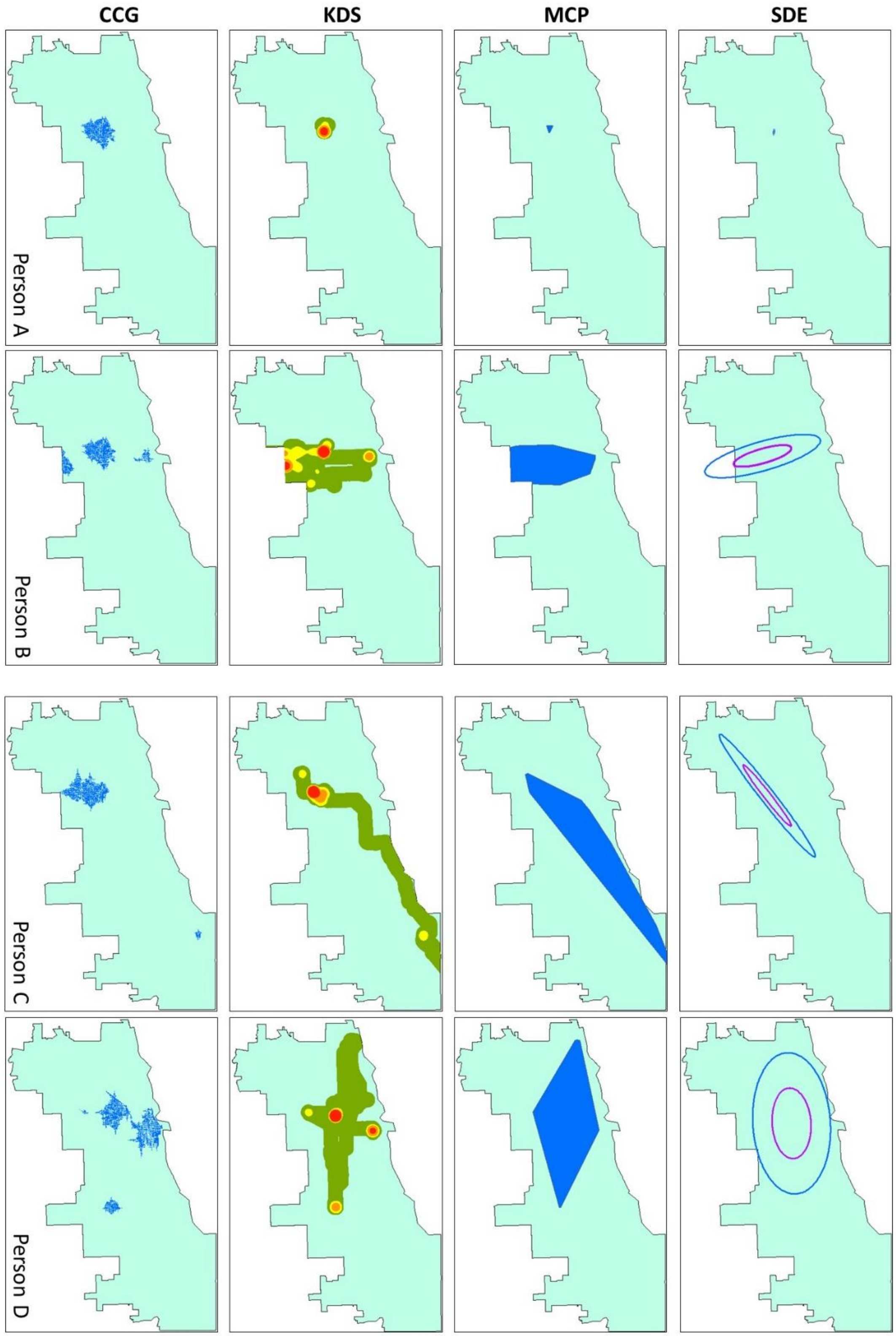


Figure 3.9 The four representative persons' activity spaces generated by different methods.

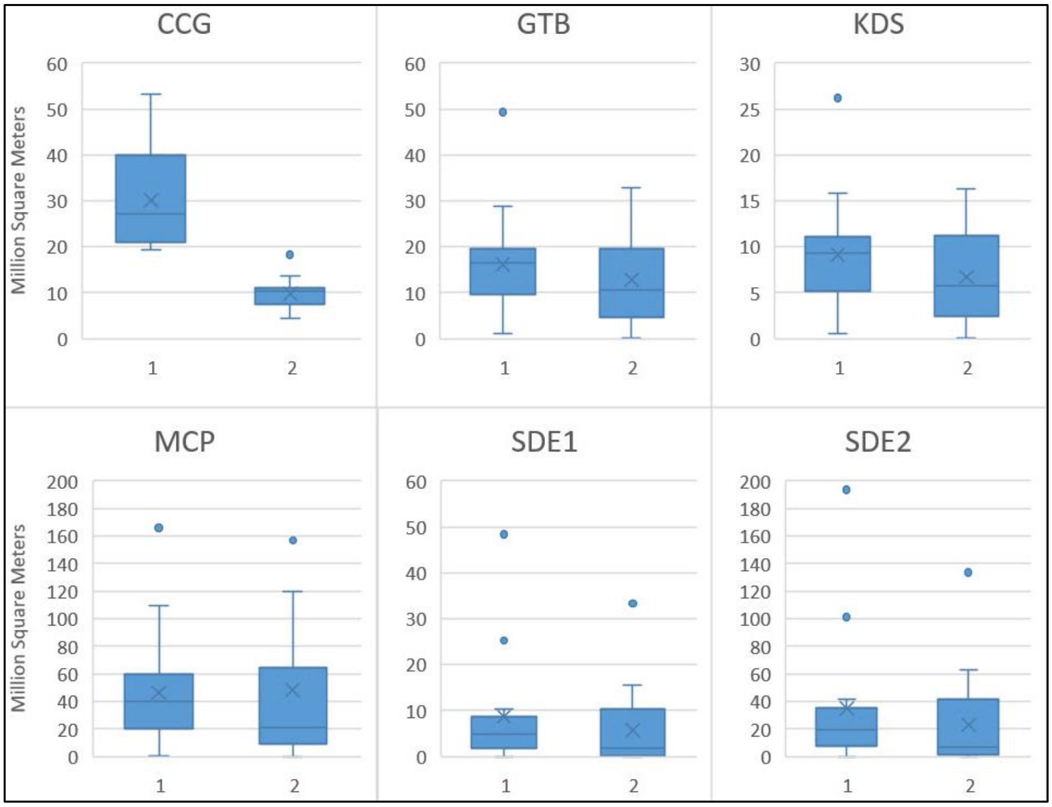


Figure 3.10 The differences in size of activity spaces for private transport (1) and public transport users (2).

3.3.2.2. Comparing physical-activity-friendly contextual exposures measured by different activity spaces

Physical-activity-friendly contextual exposures are assessed for the 31 participants by CCG activity spaces, as well as the other five methods. The assessment is based on the intersection between activity space polygon features and the physical-activity-friendly context polygon features. Contextual exposures are evaluated as the ratio of the area of the intersection polygons to the area of the activity space. In other words, they are calculated by the percentage of a participant's activity space that is physical-activity-friendly. The higher the percentage value, the higher the assessed contextual exposure. The assessment results of the 31 participants are displayed in Figure 3.11. The top section illustrates the physical-activity-friendly contextual exposures assessed by the different activity space methods, while the bottom section depicts the number of days in which the participant is physically active for 30 min or more in a typical week. As indicated in the figure, CCG-assessed contextual exposures match the number of physically active days well. For instance, Participant 1 has a high physical activity level, while CCG is the only method that yielded high contextual exposure when compared to all other methods. For Participant 14, who has a medium level of physical activity, CCG estimated a reasonable level of contextual exposure, whereas SDE gave very high levels of exposure. Again, CCG is the only method that yielded moderate exposures, while all other methods gave extremely low values for Subject 25.

To further investigate the performance of the methods for discovering the relationship between physical-activity-friendly contextual exposures assessed by the different activity space methods and physical activity level is examined using a scatter plot with trend lines (Figure 3.12). As indicated in the figure, the contextual exposures measured by CCG and KDS show a positive correlation with physical activity level, while SDE1 and SDE2 reveal a negative association. For GTB and MCP, inconsistent relationships are observed. Although both CCG and KDS reveal the positive correlation between physical-activity-friendly contextual exposures and physical activity level, CCG has more robust results based on the trend lines and plot points in the figure.

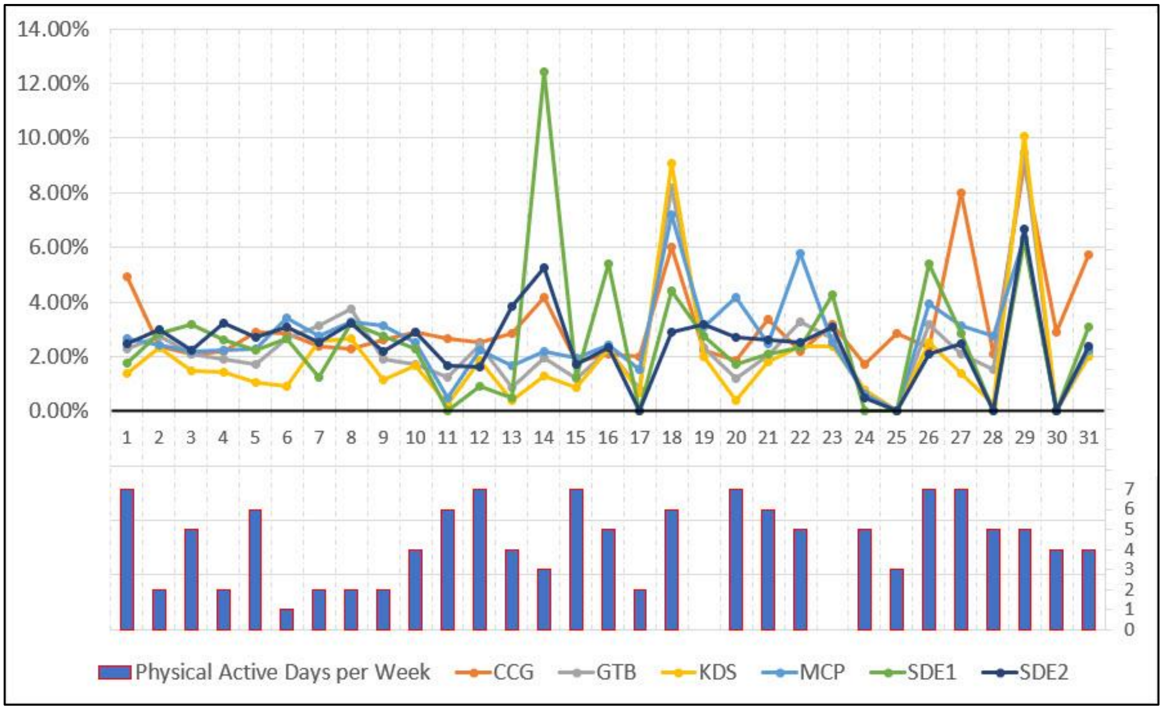


Figure 3.11 The physical-activity-friendly contextual exposures assessed by the different activity space methods and the physical active days per week for the 31 participants.

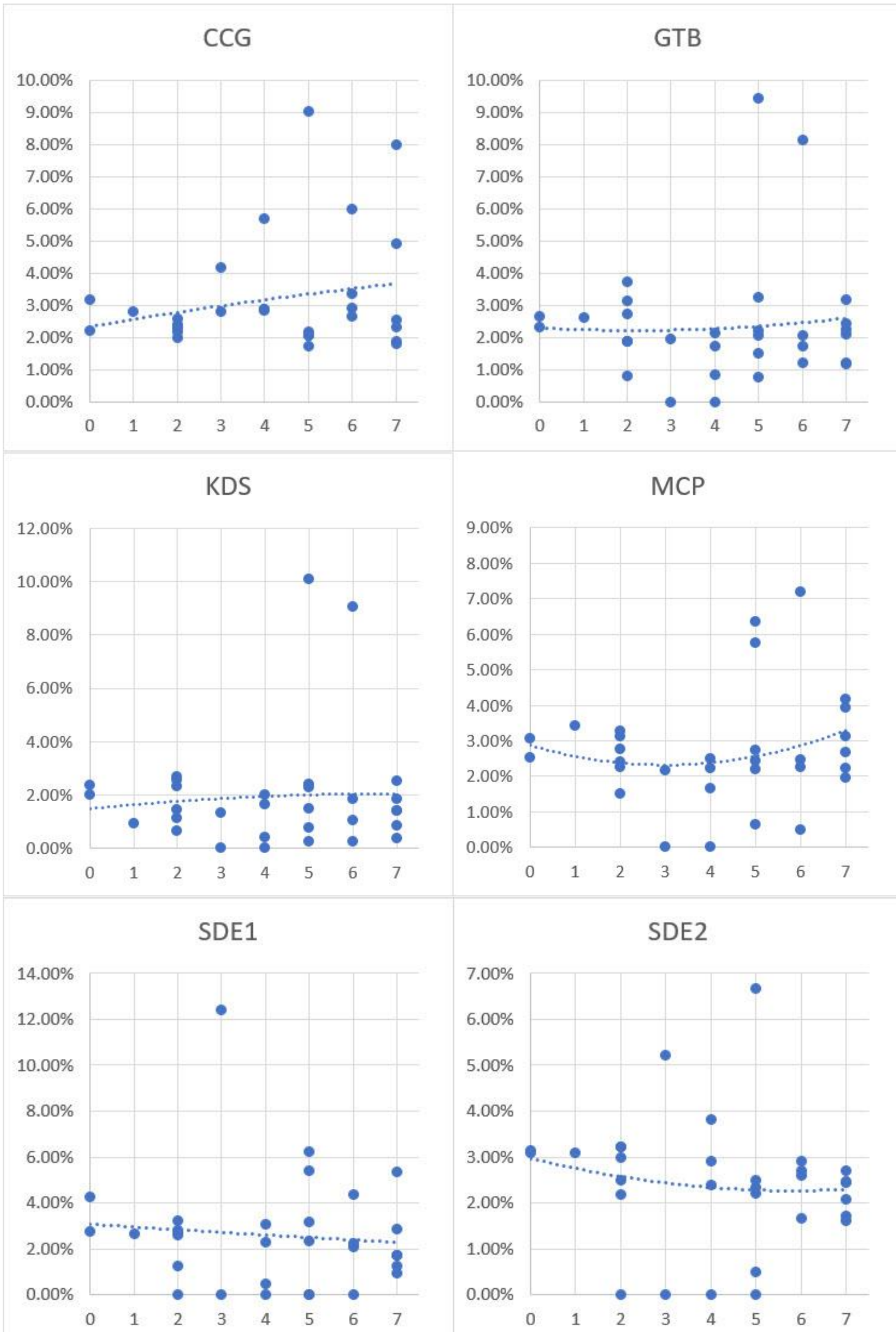


Figure 3.12 The correlation between physical-activity-friendly contextual exposures and physical activity level. (Vertical axis: the physical-activity-friendly contextual exposures level; horizontal axis: physical active days per week; trend lines are generated using the 2nd order polynomial fitting method.).

3.4. DISCUSSION

This study proposed an innovative method for delineating activity space for environmental exposure assessment based on GIS and GPS tracks data. By implementing the method with a GPS tracking dataset collected in Chicago, the proposed method generates more reasonable and reliable results when compared with other methods.

The proposed method focuses on movement pattern mining and core-area abstraction using the entire GPS trajectory patterns instead of individual GPS points. As listed in Table 3.4, the activity spaces derived with GTB, KDS, MCP are large and cover all the GPS tracking points. In these methods, each GPS point is indifferently considered as part of the activity space. The consequence of this is that the result includes not only the critical activity locations but also areas that participants passed by when traveling between activity locations. In contrast to other approaches, and by abstracting core areas based on GPS tracking data, the proposed method generates activity spaces based on the core areas that cover only the prominent parts of the points. All frequently visited activity destinations and more than 90% of the total GPS tracking records are covered in the CCG activity spaces. Further, the CCG can generate multiple areas to portrait individual activity space instead of a single continuous region, so the places that participants only passed by can be excluded. All other activity space techniques generate one activity space that unavoidably includes a large area of irrelevant space. This error can be exacerbated, as shown in Figure 3.9, when dealing with highly mobile subjects whose movement pattern is dispersed or strongly directional.

While other methods ignore the potential activity opportunities and environmental exposure around a person’s critical activity locations, which are also crucial factors for environmental exposure assessment, this study includes these potential areas to generate activity space by innovatively utilizing the crystal-growth algorithm. The idea is developed based on the fact that people are familiar with the environment and activity opportunities of the areas around their core activity locations (e.g., home, workplace). Even though they are not captured by their GPS trajectories during the tracking period, it is highly possible that people chose or will choose to undertake the daily activity and are or will be exposed to the contextual environments in these areas. For instance, in Figure 3.9, the CCG activity space of Person D contains a sub-area on the south of the city. According to the sojourn time spent in that area and the context of accessibility, an independent activity area is generated, which is centered at that core area and includes the potential activity opportunities. For other methods, GTB, KDS, and MCP activity spaces only include the actual GPS-tracked locations, while SDE doesn't even cover that area, since it is so isolated from other activity locations.

By incorporating the hexagon-grid-based accessibility-weighted plane, this is the first study that takes into account the context of accessibility for activity space delineation. Accessibility is a critical but disregarded factor in previous research. Inaccessible areas in a study area may include private houses, water bodies and restricted areas (e.g., airport). Since people can only access and undertake daily activities at accessible locations, including inaccessible areas in the activity space introduces error for delineating people's activity space as well as environmental exposure assessment. Further, the CCG method generates activity space from core areas based on specific travel time distance represented on the accessibility-weighted planes. Various levels of road networks and walkable areas are regarded as accessibility-friendly contexts with different travel speeds for private and public transport users, while people cannot trespass barrier contexts and need to bypass them, so more accurate results can be achieved by considering the context of accessibility. Additionally, the CCG method uses a hexagon grid for the weighted planes instead of the widely used raster grid to represent the context of accessibility more accurately.

The CCG method takes into account the ownership of automobiles, which is a critical factor that influences the size of people's activity space but was widely ignored in previous studies. Although living in the same neighborhood, the transportation network and facilities have different effects on the accessibility of different groups of residents. For private transport users, road networks are treated as a high-accessibility context. Thus, their activity spaces grow and are enlarged along road networks. Notice that there is a primary road passing through the area from east to west in Figure 3.8, so the activity space extends in an east-west direction. While public transport users rely heavily on the bus and metro system, their activity space extends along the bus and metro routes. Further, different from private transport users who could utilize all the road networks effectively by driving their own cars, the public transport users can only increase their travel speed by taking buses or the metro. Although the bus routes are densely distributed in the area, the average travel speed is much lower than that of private cars. Not surprisingly, the CCG activity space of public transport users is much smaller than the ones of private transport users. By considering the different environmental effects of accessibility on different residents, as indicated in Figure 3.10, the CCG activity spaces of private transport users are significantly larger than those of public transport users, which highlights the fact that residents with automobiles have high accessibility and thus a large activity space. Whereas none of the other activity space methods consider the effects of automobile ownership on activity space, which introduces error in both the delineation of activity space and the assessment of environmental exposures.

In this study, the exposure to physical-activity-friendly contexts is utilized as an example for comparing different activity space methods. Physical activity has been intensively investigated in past research, and it is highly related to obesity. The associations between physical-activity-friendly contexts and physical activity/obesity were found to be inconsistent and influenced by different delineations of activity space in environmental exposure assessment (James *et al.* 2014). For instance, Zenk et al. (Zenk *et al.* 2011) explored the environmental influences on dietary and physical activity through comparing the results generated by two different activity space methods (standard deviation ellipses and daily path areas), and inconsistent results were found. In Oliver et al.'s (Oliver *et al.* 2007) study, the influence of land use on walking behaviors was examined by using 1-km circular and line-based road network buffers. The author found that the selection of activity space methods has considerable influences on the analytical results (Oliver *et al.* 2007). In addition, different kinds of activity space methods for built environment assessment were compared by analyzing their relationship with energy balance and obesity in other studies, e.g., (James *et al.* 2014, Zhao *et al.* 2018). These results indicate that activity space delineations have a significant influence on the observed associations between environmental exposures and health outcomes or behaviors (James *et al.* 2014, Zhao *et al.* 2018). Although the results are inconsistent, many past studies have observed the positive relationship between people's exposures to physical-activity-friendly contexts and their physical activity level (Handy *et al.* 2002, Mota *et al.* 2005, Berke *et al.* 2007, Maroko *et al.* 2009, Santos *et al.* 2009, Zenk *et al.* 2011, Roberts 2017). CCG is the only activity space method that discovered such positive association in this study. The comparative results in Figures 3.11 and 3.12 support the idea that the proposed context-based crystal-growth activity space method performs better than other activity space methods in environmental exposure assessment. Although it is not feasible to use regression models to investigate the relationship between contextual exposures and participants' physical activity due to the small sample size and the lack of other independent variables, this study obtained results that reasonably justify the usefulness and effectiveness of the proposed CCG activity space method for exposure assessment and environmental health studies. Further, although this study only implements the new method for assessing environmental influences on physical activity, it can be easily applied to other environmental health studies.

In summary, the CCG activity space method achieves a balance between the actual activities captured by GPS trajectories and the context-based potential activity opportunities around core areas, which are both crucial factors for environmental exposure assessment. According to the results, the SDE and MCP are too sensitive to the pattern of the GPS trajectories; they always include much irrelevant space that makes the activity space too large. Remote activity sites that are far from the other activity locations tend to be ignored by SDE. GTB is more focused on travel behavior and therefore ignores the important activity locations. With respect to KDS, similar to all other existing methods, it disregards the context of accessibility.

As discussed above, the existing activity space methods fail to mitigate the UGCoP. Spatially, the context of accessibility is ignored, and thus the activity spaces derived indifferently include many areas that may not be accessible. Furthermore, the effects of automobile ownership on activity space are ignored, therefore introducing uncertainty in the delineation of activity space. Moreover, arbitrary cut-off distance is used for the delineation of activity space, which adds error in the assessment of environmental exposures. Temporally, the duration in which individuals experienced environmental context is treated merely as a multiplier of exposure, while the interactions with space during the time (the more time spent at the location, the more familiar with the surrounding area) is overlooked. The CCG activity space method addresses these issues and delineates individual activity space more reasonably. It thus helps mitigate the UGCoP when assessing environmental exposure by people’s activity space.

Finally, this research has several limitations that need to be explored or addressed in future studies. First, this study implements the proposed activity space method based on the 7-day GPS trajectories of a small sample of participants and compares the results with other existing methods from the aspects of the geometric characteristics of the activity space, its correspondence with subjects' daily activity, and visual interpretation. However, a larger dataset with the GPS trajectories of more subjects and further utilizing of the proposed technique for environmental exposure assessment are needed in future research to further evaluate and justify the method. Second, although the CCG is compared with other existing approaches based on GPS tracking data, there are also methods that don't rely on GPS or GIS. For instance, studies that used map-based electronic questionnaires (Chaix *et al.* 2012, Kestens *et al.* 2018), mobility surveys (Kestens *et al.* 2012) and activity space questionnaires (Shareck *et al.* 2014) also yielded useful results. Although these methods based on self-reported information may include recall bias and not be geographically accurate, they can capture more background information about participants' activities (Kestens *et al.* 2018), such as the transportation modes and social interactions (Kestens *et al.* 2017). Thus, further study is needed to compare these methods or integrate them with CCG to generate more accurate activity space. Third, the final activity space is hexagon-grid based, so distance decay functions can be applied optionally to the results to generate a weighted activity space, in which the core areas have a weigh according to their NST values. The weights of other cells can be calculated with specific distance decay functions based on travel distance from the core areas. Fourth, as the only parameter when calculating the growth extent of all core areas, for the purpose of illustration, GE_h is set to 10-mins' travel distance for all subjects in this study. However, it can also be determined based on personal mobility (such as age and health condition) to increase accuracy. Fifth, the method is based on the assumption that people are familiar with the environment and activity opportunities of the areas around the core areas, and it is highly possible that people choose to undertake the daily activity and expose to the context in these areas. However, even if only rarely, it is possible that one person may spend a lot of time at one location but is still not familiar with the surrounding area. This problem could be addressed by cross-validation with an activity diary data in future studies.

CHAPTER 4: AN ANALYTICAL FRAMEWORK FOR INTEGRATING THE SPATIOTEMPORAL DYNAMICS OF ENVIRONMENTAL CONTEXT AND INDIVIDUAL MOBILITY IN EXPOSURE ASSESSMENT[3]

4.1 INTRODUCTION

Environment-related chronic diseases are one of the biggest threats to public health. According to the World Health Organization (WHO), 22% of the global burden of disease is caused by environmental risks (Prüss-Ustün *et al.* 2016). Because environmental exposure is a significant factor that influences health behaviors and outcomes, researchers in public health and health geography have put considerable effort into assessing environmental impacts on health (Mitchell *et al.* 2016, Sallis *et al.* 2016, Browning and Lee 2017). Evidence shows that exposures to different environmental factors, such as air and noise pollution (Eriksson *et al.* 2007, Ta *et al.* 2015, Park and Kwan 2017), the built environment (Troped *et al.* 2010, Ding *et al.* 2011, Sallis *et al.* 2016, Browning and Lee 2017) and the food environment (Morland and Evenson 2009, Caspi *et al.* 2012, Cobb *et al.* 2015, Gamba *et al.* 2015, Lytle and Sokol 2017), have significant associations with various health behaviors, including physical activity (Lachowycz *et al.* 2012, Koohsari *et al.* 2015, Sallis *et al.* 2016), tobacco and drug use (Kwan *et al.* 2011, Epstein *et al.* 2014, Lipperman-Kreda *et al.* 2015, Shareck *et al.* 2015), and health outcomes, which includes obesity and obesity-related disease (Andersen *et al.* 2008, Oliver and Hayes 2008, Chaix 2009, Millstein *et al.* 2009, Seliske *et al.* 2009) and mental health disorders (Curtis 2010, Stigsdotter *et al.* 2010, Houle and Light 2014, Wheaton and Clarke 2016).

---

[3] Reprint, with permission, from Wang, J. and Kwan, M.-P., 2018. An Analytical Framework for Integrating the Spatiotemporal Dynamics of Environmental Context and Individual Mobility in Exposure Assessment : A Study on the Relationship between Food Environment Exposures and Body Weight. International Journal of Environmental Research and Public Health, 15 (9), 2022–2045.

The food environment, among different environmental factors, is one of the critical factors that could lead to obesity (Kestens *et al.* 2012) and other obesity-related chronic diseases such as type II diabetes (Auchincloss 2009) and cardiovascular diseases (PAGAC 2008), whose prevalence has increased rapidly in recent decades (Ogden *et al.* 2014). Increasing public concerns have prompted a growing number of studies on the effects of food environment exposures on obesity. Previous studies found that living in food desserts (Walker *et al.* 2010) and exposures to unhealthy food outlets (Inagami *et al.* 2009) may encourage unhealthy food intake behavior that is associated with a higher likelihood of obesity. Furthermore, the association between food environment exposures and obesity has been utilized in developing intervention strategies to improve public health by numerous institutions worldwide (Reisig and Hobbiss 2000, Committee on Accelerating Progress in Obesity 2012, Vandevijvere *et al.* 2015).

However, the findings of the effects of food environment exposures on obesity are inconsistent (Holsten 2009, Chen and Kwan 2015, Cobb *et al.* 2015). Although a higher likelihood of obesity has been found to be significantly associated with exposures to unhealthy food (e.g., fast food restaurants) in many studies (Maddock 2004, Morland and Evenson 2009), it was not observed in other research (Zick *et al.* 2009, Jilcott *et al.* 2011, Lee 2012). For example, exposures to fast food restaurants were found to be positively associated with the prevalence of obesity in some studies (Davis and Carpenter 2009, Inagami *et al.* 2009, Li *et al.* 2009), while no correlation (Jeffery *et al.* 2006, Dunn *et al.* 2012) or even negative association (Black *et al.* 2010) was observed in other studies. The inconsistent findings regarding the effects of the food environment on obesity bring enormous challenges on implementing effective policy to improve public health.

Many potential issues could cause these inconsistent findings, including the modifiable areal unit problem (Kwan 2018b), the uncertain geographic context problem (Kwan 2018b), and spatial non-stationarity (Wang, Lee, *et al.* 2018). To examine whether the food environment has significant influences on obesity, an important task is to accurately measure individual exposure to relevant environmental factors (Kwan *et al.* 2018). Past studies that examine the effects of environmental exposures on health outcomes have predominantly used residential neighborhoods as contextual units (Frank *et al.* 2005). In these studies, residential neighborhoods were defined either by the administration units (e.g., census tracts) in which people's homes are located or by buffer areas with a specific radius around people's home location (Feng *et al.* 2010, Leal and Chaix 2011, Clark and Scott 2014). However, residential neighborhoods only partially represent the environmental context that affects people's health, since people move around in their daily life (Wiehe, Hoch, *et al.* 2008, Kwan 2009a, Basta *et al.* 2010). Identifying environmental context based solely on the residential neighborhood may thus lead to inaccurate contextual exposure assessment and erroneous results concerning the relationships between environmental contexts and health outcomes (Cummins 2007, Kwan 2013). This methodological issue may contribute to the inconsistent findings of past studies and has been articulated as the uncertain geographic context problem (UGCoP) by Kwan (2012a).

The UGCoP refers to the problem that findings about the effects of environmental factors (e.g., exposure to fast food restaurants) on individual health outcomes (e.g., obesity) could be affected by how contextual units are geographically delineated and the "temporal uncertainty in the timing and duration in which individuals experienced these contextual influences" (Kwan 2012b). In light of the dynamic nature of people's daily activity, people's movement in space and time should be taken into account while measuring their food environment exposures and its effects on obesity (Kwan 2013). To mitigate the UGCoP, portable GPS devices can be utilized to accurately trace human movement in space and time, and advanced GIS methods can be used to relate activity locations to relevant environmental risk factors (Almanza *et al.* 2012, Kwan 2012b). Further, GPS trajectories can be used to derive human activity space, which is more representative of people's daily context than the residential neighborhood (Chaix, Méline, Duncan, Merrien, *et al.* 2013). The integration of GPS and GIS provides a powerful means for investigating the relationships between environmental contexts and health outcomes (Wiehe, Hoch, *et al.* 2008, Maddison and Ni Mhurchu 2009).

Individual GPS trajectories capture people's movement in space and time and thus help mitigate the UGCoP through more accurately delineating contextual units. A growing number of studies have started to adopt GPS-based activity space methods to investigate environmental effects on health outcomes (Kestens *et al.* 2012, Laatikainen *et al.* 2018, Wang, Kwan, *et al.* 2018). However, environmental contexts can vary over time in a highly complex manner and are thus temporally uncertain (Kwan 2018b). Some contextual variables change over the 24-hour period of a day (e.g., air pollution and the food environment), and some change over the seasons (Gulliver and Briggs 2005, Park and Kwan 2017). The temporal uncertainty of the environmental context is mostly ignored and not taken into account in environmental exposure assessment in previous studies. With regard to the food environment, food outlets open and close according to their daily schedules. Furthermore, many food outlets have different opening hours for weekdays and weekends. Given the complexity and spatiotemporal uncertainties of the food environment, it is highly challenging to accurately delineate the environmental context and assess individual exposures to the food environment. The UGCoP may introduce considerable errors to the results if the spatial and temporal variability of the food environment and people's daily movement in relation to such environment are not appropriately considered, but most previous research has paid little attention to them.

To address the challenges of the UGCoP, this paper proposes an analytical framework that utilizes GIS, time geography and GPS trajectory data to assess individual food environmental exposures for environmental health studies. In the framework, an environmental context cube (ECC) integrates variations in the food environment over space and time into a 3-D cube. By buffering individual GPS trajectories in 3-D to generate the individual space-time tunnel (ISTT) and projecting it into the ECC, environmental exposures can be derived and assessed by identifying the 3-D intersection of the ISTT and ECC. Based on the intersection, we calculate the environmental context exposure index (ECEI) as a standardized measure of individual exposure to the food environment. Considering both the spatiotemporal variations in the food environment and the dynamics of people's daily movement, the ECEI may provide a more accurate and reliable measurement of individual exposure to the food environment. The ECEI is utilized in this study to explore the relationship between individual food environment exposures and the overweight status for 46 participants using data collected with GPS in Columbus, Ohio, and binary logistic regression models. The results indicate that the proposed framework is effective for assessing individual exposures and investigating their health effects when compared with other widely used food environment exposures assessment methods. Addressing the spatiotemporal variations of contextual influences, the framework may help mitigate the spatial and temporal uncertainties in the food environment in public health studies. Further, the methodology is also useful in a wide range of environmental health research.

4.2 THE PROPOSED ANALYTICAL FRAMEWORK

This study proposes an analytical framework for assessing the effects of environmental exposures on individual health outcomes using the food environment as an example. The framework seeks to more accurately measure individual food environment exposures so that more robust research findings can be obtained. Figure 4.1 illustrates the framework that integrates GIS, time geography, and GPS tracking. For this study, the food environment (BMI-unhealthy food outlets) is selected to generate 3-D environmental context cubes (ECC) for weekdays, Saturdays, and Sundays, which represent the spatial and temporal dynamics of the environmental contexts. By buffering individual GPS trajectories in 3-D, an individual space-time tunnel (ISTT) is generated to represent individual exposure space. By projecting the 3-D ISTT into the corresponding 3-D ECC and identifying their intersection, the individual environmental context exposure index (ECEI) can be derived as a standardized exposure measure. Based on the ECEI, the effect of the food environment on overweight is explored with statistical models, and the results are compared with other methods. Details of the dataset, the ECC, the ISTT, and the ECEI are discussed in the following sections.

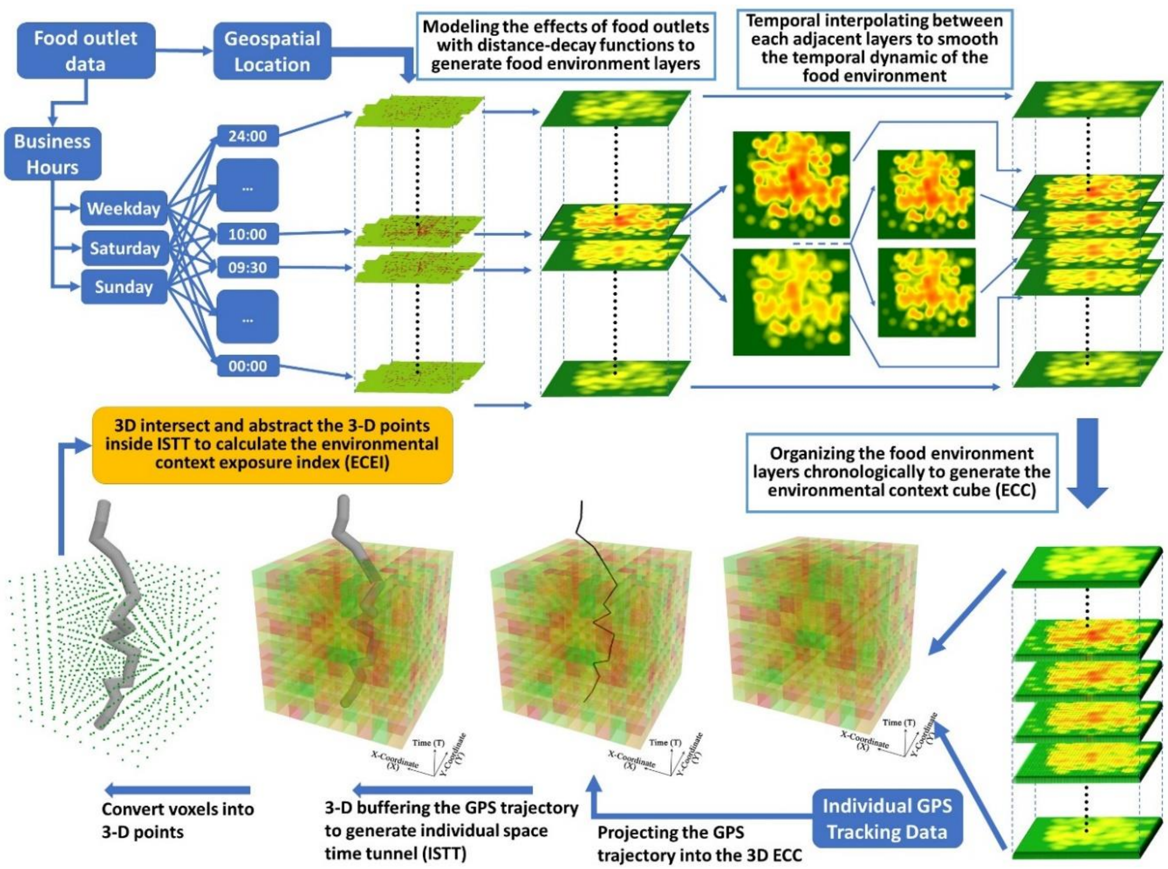


Figure 4.1 The proposed analytical framework.

4.2.1 Study Area and Data

The study area for this research is Franklin County (Ohio, U.S.), which is part of the Columbus metropolitan area and where the city of Columbus is located. It is the second-most-populated county in Ohio where the percentage of obese or overweight adults is 63.9% (Franklin County Community Health Needs Assessment Steering Committee 2013). The county includes urban, suburban, and some rural areas. This characteristic is helpful for a study that seeks to consider the influences of various land uses on people's health outcomes. In addition, Franklin County has a diverse racial composition, and also has both wealthy and impoverished communities. Further, there are 3727 food outlets in the county that include many kinds of food retailers (e.g., fast food restaurants, full-service restaurants, and supermarkets) that provide different types of foods. These features of Franklin County facilitate the identification of the spatial heterogeneity of the food environment and the differences in the levels of exposure among the county's population with various socio-economic statuses.

The GPS trajectory dataset used in this research was collected as part of a larger study that examines the influence of parks on people's physical activity in four U.S. cities: Albuquerque (NM), Chapel Hill and Durham (NC), Columbus (OH), and Philadelphia (PA). In each of these cities, participants were recruited in person in selected public parks and neighborhoods surrounding these parks following household interviews. For each selected park and its surrounding neighborhoods, about 300 persons from different socio-economic backgrounds were randomly solicited to participate in the study. In the end, 238 subjects participated in the study and 51 of them were from the Columbus study site. Participants in the study were asked to wear a GPS and an accelerometer for three consecutive weeks. The data were collected from August 2009 to October 2010 in three selected seasons (spring, summer, and fall) to avoid the winter months in which people may undertake fewer outdoor physical activities (since the purpose of the larger project was to assess the influence of parks on people's physical activity). Geographic location was recorded by the GPS devices with a time interval of one minute. In addition, data about subjects' demographic, anthropometric and socio-economic statuses were also collected. Subjects' overweight status in the dataset was assessed by the Body Mass Index (BMI), which was calculated by dividing the subject's weight (kg) with his or her height in meters squared (m^2).

The environmental context data for this study were derived from a comprehensive digital geographic database of Franklin County maintained by the Franklin County Auditor's Office. It includes the attributes and physical boundaries of relevant environmental contexts. Food outlets data were derived from the food license data of Franklin County, and their business hours were collected and confirmed using Google Map and phone calls. These data include each food outlet's business name, geographic location, business hours and business category according to standard industrial classifications.

4.2.2 Data Preprocessing

GPS signals may be absent in locations near tall buildings or under dense tree canopies, and this may lead to gaps in GPS tracking data. Data preprocessing is thus necessary to improve the reliability and usefulness of GPS data. Consistent with the procedures used by Wiehe et al. (2008), missing GPS records in the Columbus GPS dataset were inserted at the location of the earlier point if the distance between two temporally adjacent records bounding a period of missing data was less than 30 m. If this distance was longer than 30 m and the gap between two recorded GPS points was less than 1 hour, interim 1-minute time points were imputed. Missing GPS points for time periods longer than an hour were considered missing and not imputed. Further, only survey days with eight or more hours of valid GPS records for each subject were included in the analysis.

Although the original dataset had 51 participants, only subjects with valid GPS records for at least five weekdays and two weekend days were included in the analysis. This ensured that there were sufficient data for the selected subjects for at least seven days that covered their daily activities in both weekdays and weekend days. As a result, 46 participants were finally selected as valid subjects for further analysis. Although the sample size is not large, the subjects cover a range of socio-economic attributes (e.g., age and education level) and are thus useful for this exploratory study of the proposed analytical framework for food environment exposure measurement.

Table 4.1 shows the demographic and socio-economic characteristics of the 46 participants in the sample used in the study. These participants are predominately female (60.87%) and younger people. All of them are adults, and only 2.18% are seniors older than 65. With respect to the education level, 56.52% of the subjects have a college degree or higher, while 43.58% of them have a high school degree or lower. The overweight status among the 46 participants is balanced in that half of them are overweight, and the other half are not overweight.

Table 4.1 The socio-demographic characteristics of the participants.

Socio-demographic Variables		**Percentage**
Gender	Male	39.13%
	Female	60.87%
Age (years old)	18–30	56.52%
	31–65	41.30%
	65 +	2.18%
Education	With College Degree or Higher (≥ College Degree)	56.52%
	With High School Degree or Lower (< College Degree)	43.48%
Overweight Status	Overweight	50%
	Non-overweight	50%

4.2.3 Representing Dynamic Environmental Contexts Using the Environmental Context Cube

Any attempt to measure exposures to the food environment needs to begin with representing the food environment, which in turn requires researchers to consider the location and distribution of food outlets as well as the kinds of food those outlets provide. The geographic range of influence from food outlets can be assessed by creating homogeneous buffer areas covering food outlet locations with a specific distance (such as 100 m or 1 km). Importantly, however, representation of food outlets' effects on health behavior should take into account the effect of distance decay rather than using arbitrary distance cutoffs: Environmental effects change as a function of distance, with locations farther from a food outlet less influenced by that outlet than nearer locations are. A few researchers have included distance-decay functions (Páez *et al.* 2010, Dai and Wang 2011, Kestens *et al.* 2012, Lamichhane *et al.* 2012, Lee *et al.* 2013, Moore *et al.* 2013, Xu *et al.* 2015) as part of their food environment studies with a view to accounting for the effect of distance, but even these studies have treated the food environment statically and have failed to consider the dynamic features of the food environment (e.g., food outlets' opening and closing at different times of the day).

As some researchers have noted (Kwan 2012a, Chen and Kwan 2015), environmental contexts undergo continuous change. For example, air pollution levels differ throughout the day. For this reason, exposure assessment may produce erroneous results if the variability of the environmental context is ignored (Park and Kwan 2017). Furthermore, the contextual influences of the food environment may also differ with time of day. Previous studies have largely ignored temporal variations in the food environment, although most food outlets operate on specific schedules and offer their services only during certain hours. Some even feature different schedules for weekdays and weekends. Accordingly, Chen and Clark (2016), arguing that space-only methods of exposure assessment overlook the dynamic features of the food environment, proposed a spatiotemporal method that takes into account food outlets' business hours when measuring access to food retailers. Although portrayal of the food environment spatiotemporally in Chen and Clark's (2016) study is an important step forward for the study of environmental health, the study measures food access using census tracts as its contextual units and thus may still be susceptible to the UGCoP. As a result, further development is needed to more accurately assess individual exposures to the food environment and to mitigate the UGCoP.

The ECC is developed in this study to address both spatial and temporal uncertainties in the food environment while accurately assessing individual exposures to that environment. It is designed to capture the complex dynamics of environmental contexts as well as individual exposures. Indeed, this extension of the space-time cube (Kwan 2000) is explicitly designed for use in environmental health research. The base of the space-time cube (or space-time aquarium), which is a time-geographic construct first introduced by Hägerstrand in the 1960s, represents the geographic contexts of the study area (*x*-axis and *y*-axis), with 3-D lines inside the cube representing an individual's movement trajectories. The cube's vertical dimension (*z*-axis) represents time. Figure 4.2 in the Appendix illustrates a space-time cube that integrates geographic contexts and GPS trajectories. Note, however, that this representation visualizes only individual movement trajectories in 3-D, whereas the environmental context is represented on a 2-D plane. Constrained by the 2-D plane, the environmental context can be visualized and analyzed at only one time point using this space-time cube framework. In real-world contexts, however, environmental contexts and their influence on moving subjects may change over both space and time in highly complex ways. Representation of the environmental context should thus be also extended to capture and represent the dynamic features of the environment by integrating time as the third dimension.

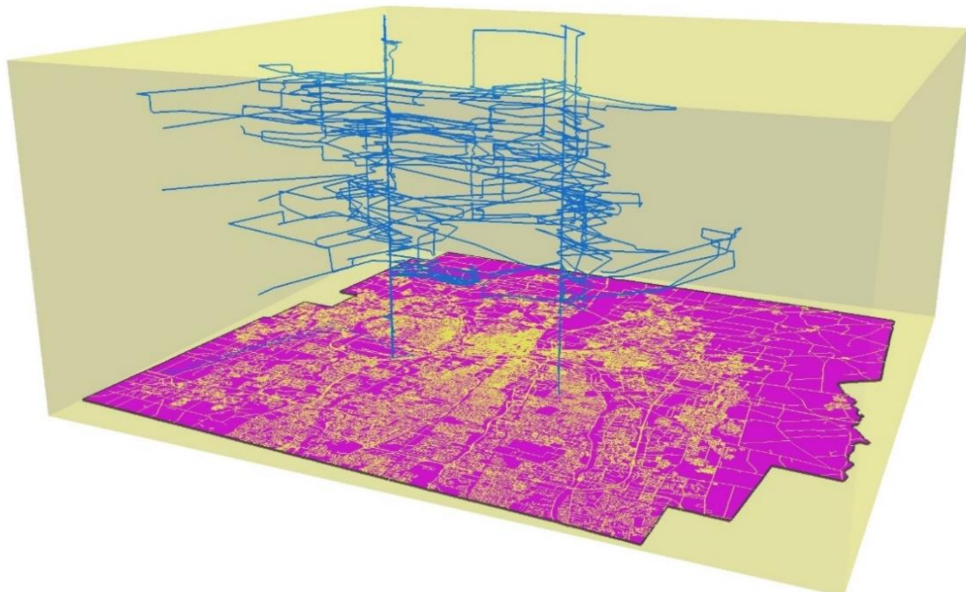


Figure 4.2 An example of the space-time cube (aquarium).

By extending the traditional space-time cube, we propose the environmental context cube (ECC) as a new analytical framework for analyzing people's movement and their dynamic relationships with their environmental context (e.g., the food environment). The ECC is a collection of 3-D voxels arranged on a regular grid in 3-D space. The value of each voxel represents the environmental context at a specific geographic location (*x*- and *y*-coordinate) at a specific time (*z*-coordinate). Thus, spatial and temporal variations in the environmental context are rendered as the different values of the voxels in 3-D space at various locations and times. In the temporal dimension of the cube, layers of voxels constitute the ECC, with each layer representing the spatial configuration of the environmental context at a specific time of day. The size of the voxels in the 3-D space represents the spatial and temporal resolutions: The higher the spatial resolution, the more detailed the spatial variations represented; and the finer the temporal resolution, the more detailed is the representation of the temporal dynamics. Different spatiotemporal resolutions of the

ECC may affect the accuracy with which variations in the environmental context can be captured. To thoroughly examine the ECC while exploring the effect of spatiotemporal resolutions (or scale effects) on measurement accuracy, we built ECCs featuring combinations of three different spatial resolutions (100 m × 100 m, 150 m × 150 m, 200 m × 200 m) and two temporal resolutions (30 min, 10 min), and then evaluated and compared their performance.

In this manner, we constructed a series of 3-D ECCs to represent the unhealthy food environment of the study area. We classified the food outlets in the study area into three categories as described by Rundle et al. (2009)—BMI-healthy, BMI-unhealthy, BMI-neutral. Because exposure to BMI-unhealthy food outlets may be related to high BMI, as has been observed in many previous studies (Maddock 2004, Davis and Carpenter 2009, Li *et al.* 2009, Mellor *et al.* 2011), BMI-unhealthy food outlets were used in this study to generate 3-D ECCs for measuring individual exposure to unhealthy food environment. In Rundle et al.'s (2009) classification system, BMI-unhealthy food outlets include fast-food restaurants, convenience stores, meat markets, pizzerias, bakeries, and candy and nut stores. Franklin County, the study area, contains 1,645 BMI-unhealthy food outlets. All of them are included in the food environment analysis in this study. Figure 4.3 shows the location of these BMI-unhealthy food outlets.

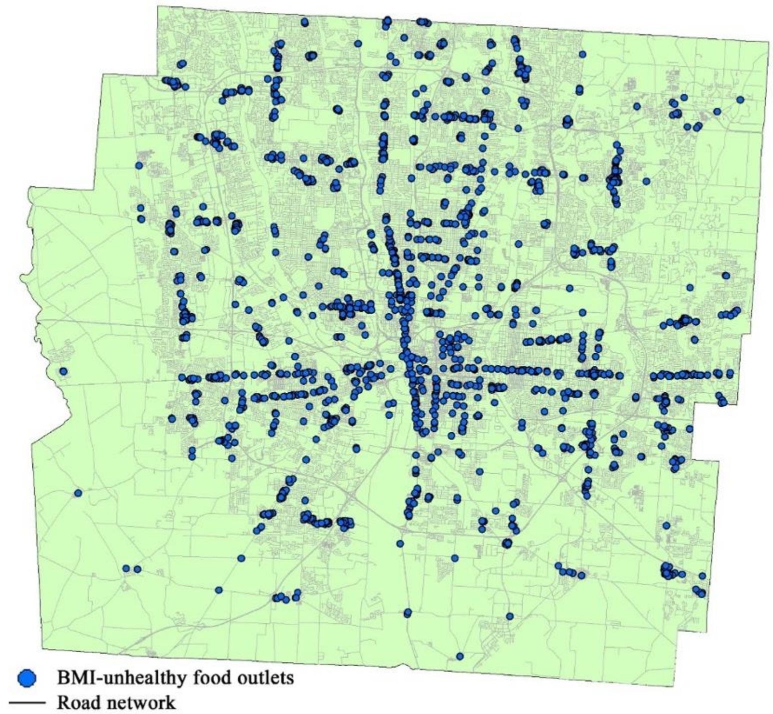


Figure 4.3 The study area (Franklin County) and the location of BMI-unhealthy food outlets.

In analyzing the business hours of the 1645 BMI-unhealthy food outlets included in the study, we found that many food outlets had different schedules on weekdays and weekends; some even had different schedules on Saturdays and Sundays. To give just one example, a restaurant might be open from 10 a.m. to 10 p.m. on weekdays, from 11 a.m. to 10 p.m. on Saturdays, and from 11 a.m. to 8 p.m. on Sundays. Accordingly, we generated three different environmental context cubes (ECCs) to better represent this dynamic food environment: one for weekdays, one for Saturdays, and the other for Sundays. To construct the ECC for one day, we first generated layers of the food environment at different times of the day. For ECCs with a temporal resolution of 30 min, we generated a food environment layer for each of the 48 half-hour time slots in the day. For each time slot, a food environment layer was created to estimate the extent and degree of environmental effects of BMI-unhealthy food outlets based on the locations of the outlets open at that specific time.

As already noted, any assessment of the environmental influence of food outlets should take into account the effects of distance decay. For this reason, based on the location of the food outlets operating during each of the 48 time slots of a day, we modeled the decline in each food outlet's influence using three distance-decay methods that were used in previous environmental health studies (Páez *et al.* 2010, Dai and Wang 2011, Kestens *et al.* 2012, Lamichhane *et al.* 2012, Lee *et al.* 2013, Moore *et al.* 2013, Xu *et al.* 2015): kernel density estimation (KD), an inverse-square distance-decay function (ISDD), and a negative-exponential distance-decay function (NEDD). The food environment at a specific time of day (a time slot) was thus represented as a raster layer created by estimating the extent and degree of the environmental effects of BMI-unhealthy food outlets based on the locations of food outlets operating at that time of day, using one of the three distance-decay functions. Figure 4.4 shows the food environment layers at three different time slots of the same day, as generated by the three distance-decay methods. Separate food environment layers were generated for the 48 time slots using each of these methods for three kinds of days (weekdays, Saturdays and Sundays). Then, for each kind of day, the 48 food environment raster layers were voxelized with the unit size of 30 min and mapped to the *z*-axis. In this way, a 3-D environmental context cube (ECC) was constructed by organizing the 48 voxelized layers chronologically using the assigned z-values (which represent the specific time corresponding to each layer). Because the ECCs were implemented using three distance-decay methods (KD, ISDD, NEDD) in three spatial resolutions (100 m × 100 m, 150 m × 150 m, 200 m × 200 m) for three kinds of day (weekdays, Saturdays, and Sundays), 27 ECCs were ultimately constructed with a temporal resolution of 30 min.

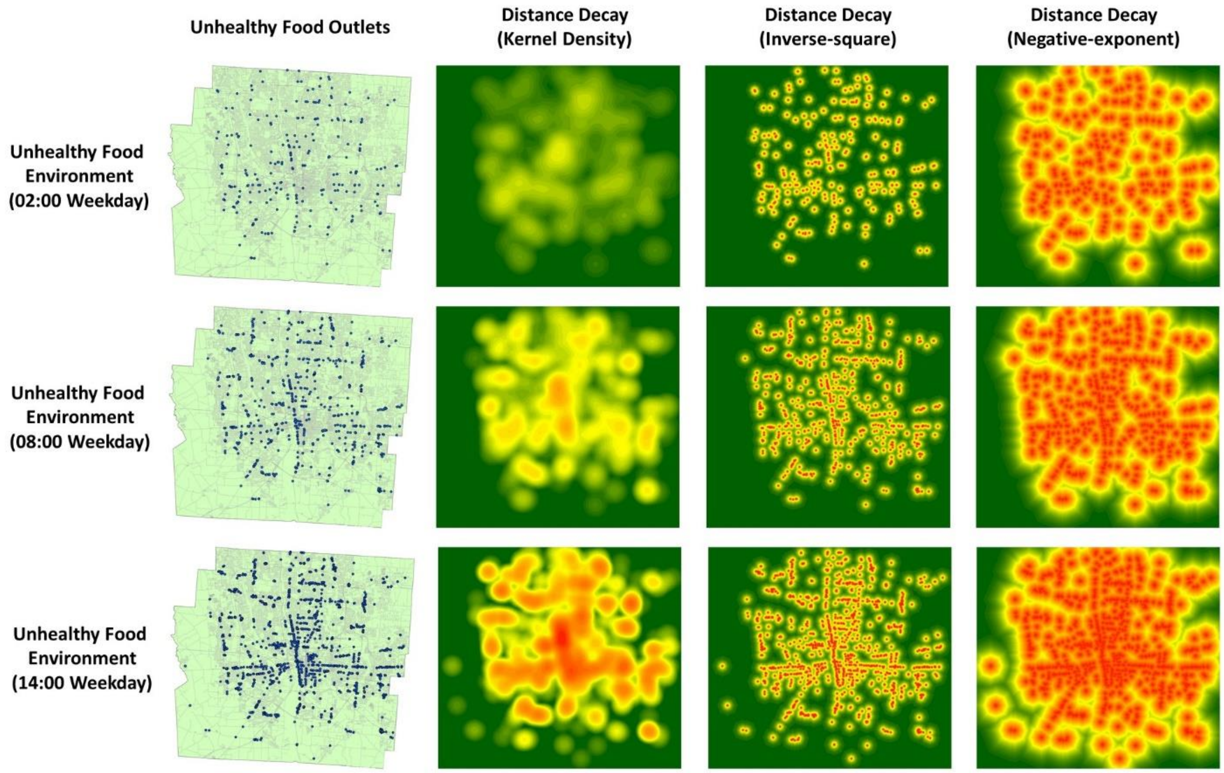


Figure 4.4 The food environment layers at different time of a day generated by the three different distance-decay methods (the food environment layers with spatial resolution of 100 m × 100 m are used in this figure to illustrate the three distance-decay methods).

To capture more details of the temporal variations in the food environment, another 27 environmental context cubes (ECCs) with a temporal resolution of 10 min were constructed using the same three distance-decay methods, three spatial resolutions, and three kinds of day. Each of these ECCs has 144 food environment layers, where each layer represents each of the 10-minute slots that make up the 24 hours of a day (e.g., 10:10 a.m., 10:20 a.m. and so on). These layers were calculated by raster algebra using linear interpolation, in which the pixel value of an additional layer (in addition to the original 48 layers at the resolution of 30 min) was calculated using linear polynomials based on the values of the corresponding pixels in the two temporally adjacent layers among the 48 30-minute resolution layers. These 144 food environment layers were then voxelized, mapped to the *z*-axis, and organized chronologically to form the ECCs with a temporal resolution of 10 min. This higher temporal resolution allowed the temporal dynamics of the food environment in any particular day to be better represented. Figure 4.5 illustrates the process of the temporal interpolation as well as the generation of a 3-D ECC and its food environment layers.

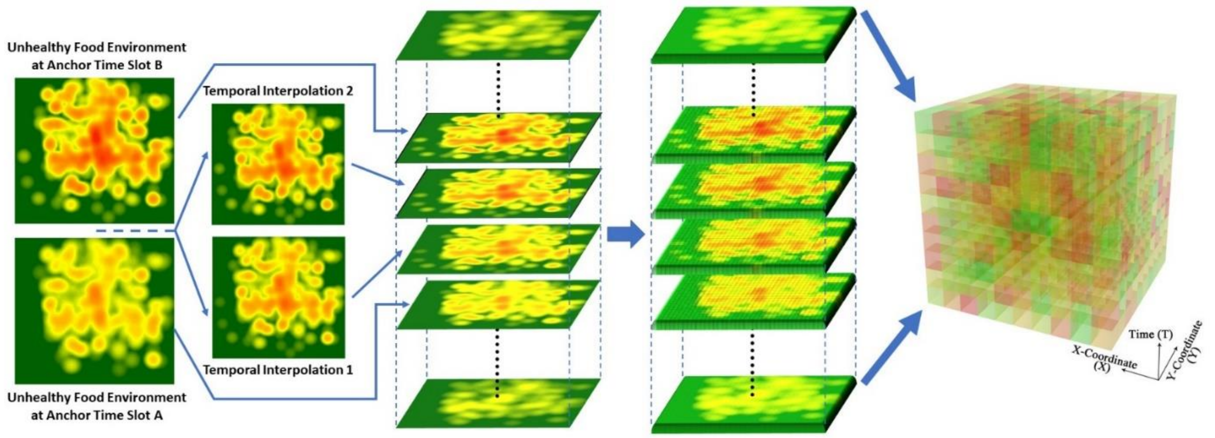


Figure 4.5 The process of the temporal interpolation and the generation of the 3-D environmental context cube with food environment layers.

To facilitate the implementation and computation of the 3-D environmental context cubes (ECCs), we converted each 3-D ECC to a 3-D point cloud, with each voxel in the cube represented by a point in the cloud at the centroid of the original voxel. As shown in Figure 4.6, the three dimensions of the points correspond to the *x*-coordinate (X), *y*-coordinate (Y), and time (T) seen in the original ECC. The values of the environmental factors were stored as an attribute table linked to each point in the 3-D point cloud.

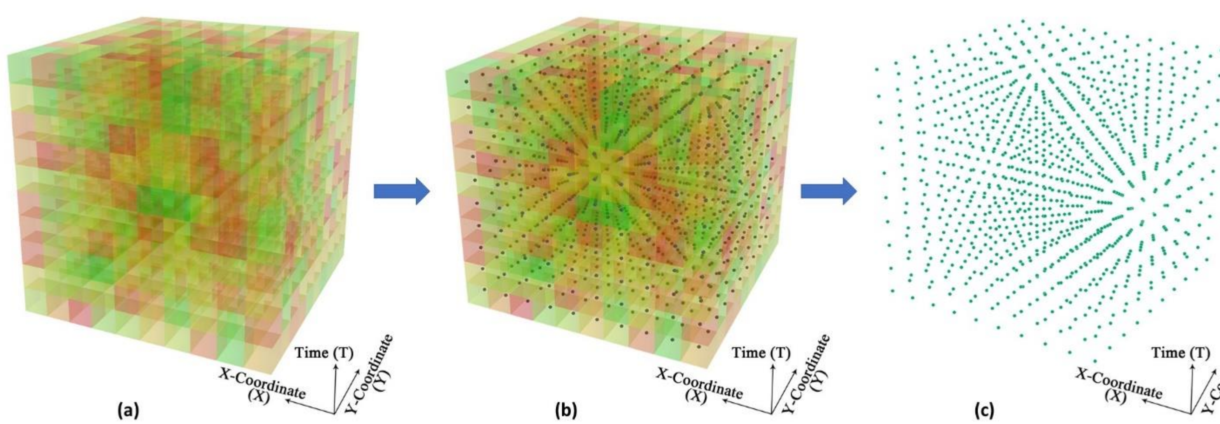


Figure 4.6 Implementation of the 3-D environmental context cube using a 3-D point cloud. (**a**): an environmental context cube, (**b**) voxels in the cube represented by points at the centroid of the original voxels, (**c**) the corresponding 3-D point cloud.

4.2.4 Capturing the Spatiotemporal Exposure Space with Individual Space-Time Tunnel

Using GPS trajectory data for individual exposure assessment could capture the daily movement of people and thus help mitigate the UGCoP to some degree. However, existing methods capture only the spatial extents of individuals' activities using 2-D polygons that represent a person's activity space (e.g., GPS trajectory buffers, standard deviation ellipses, and minimum convex polygons) or, at most, weigh the accumulated time spent at different activity locations (e.g., kernel density surface, context-based crystal-growth activity space (Wang, Kwan, *et al.* 2018)), where environmental contexts were considered statically and the dynamics of the food environment ignored. Although these methods consider the accumulated time that an individual spends at different locations (e.g., person A spends 8 hours at the workplace on weekdays), they ignore temporal variations in people's location of activity (e.g., person A stays at the workplace from 8 a.m. to 12 p.m. and from 1 p.m. to 5 p.m. on weekdays). Knowledge of the exact times when a person is at a location is essential for understanding the resulting level of exposure to the food environment. For example, consider a person who visits an area where many fast food restaurants may be found but does so at 1 a.m., when they are closed. Existing activity space methods would include this occurrence in the environmental context exposure assessment even though the person was not actually exposed to fast food restaurants at that time. In this way, overlooking temporal variations in the food context and the exact times when people undertake their daily activities at various locations may introduce measurement error.

To help address the temporal uncertainty that is an essential element of the environmental context and the dynamics of people's daily activity, we propose the individual space-time tunnel (ISTT) as a way of representing the individual exposure space. The ISTT was generated by a 3-D buffer of an individual's GPS trajectory at a specific distance (e.g., 100 m) in a 3-D space. To generate the ISTT, GPS trajectories were projected into the ECC according to the geographic coordinates and timestamps of the GPS records. As shown in Figure 4.7a, the GPS trajectory of a subject is projected into the 3-D space of an ECC, much like the space-time paths inside a space-time aquarium. The voxels along a particular trajectory and its surrounding areas constitute the environmental context that influences the corresponding subject. Thus, environmental exposure should be derived using a 3-D buffer space of appropriate radius around people's movement trajectories, as shown in Figure 4.7b. The buffer radius B_r, a user-defined parameter that represents the effective range of a particular environmental influence, can vary for different population groups based on individual socio-demographic attributes (e.g., age). For example, older adults and children may have a smaller B_r than adolescents do, because they both tend to have lower mobility. We might also define B_r as a function of travel velocity. For example, we might associate higher velocity with smaller B_r, noting that higher velocity (i.e., quicker bypass) may allow for less influence on the subject from the environmental context around a location. For purpose of illustration, we set B_r to 100 m in this exploratory study.

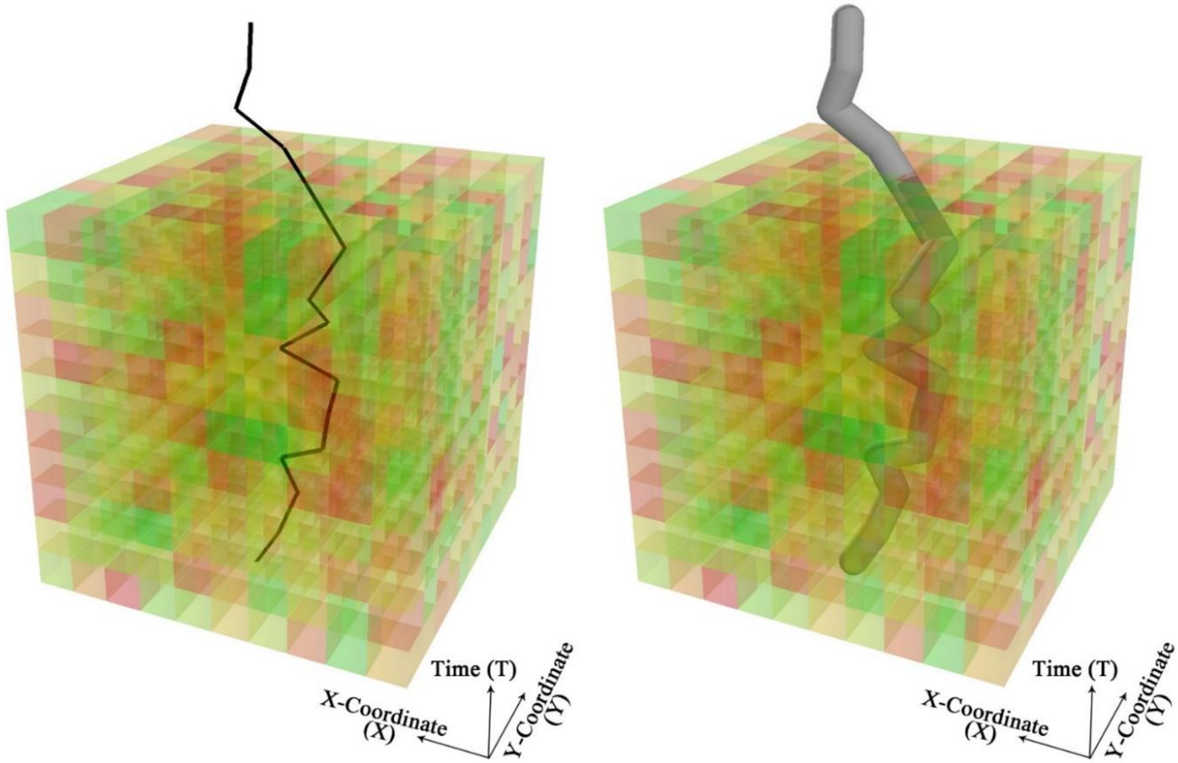


Figure 4.7 A GPS trajectory (**a**) and an individual space-time tunnel (**b**) projected into an environmental context cube.

4.2.5 Measuring Food Environment Exposure with the Environmental Context Exposure Index

The proposed 3-D environmental context cube (ECC) can capture the complexity and dynamics of the food environment, and the individual space-time tunnel (ISTT) can delineate the individual spatiotemporal exposure space by integrating spatial as well as temporal variations in a person's daily activities. As a result, individual exposures to the food environment can be derived by the 3-D intersection of the ECC and the ISTT: By projecting the 3-D ISTT into the corresponding point cloud of the ECC, as illustrated in Figure 4.8, we can link exposure to the food environment with all points located inside the ISTT in the 3-D space. The results of the 3-D intersection allow calculation of the environmental context exposure index (ECEI), which in turn allows the measurement of individual exposures to the environmental context. By capturing the extent to which a person is exposed to the relevant environmental context during each time unit throughout a day, the ECEI provides a new method for quantifying an individual's level of exposures to the food environment. In this study, the ECEI was used to analyze the association between individual exposure to the food environment and health outcomes.

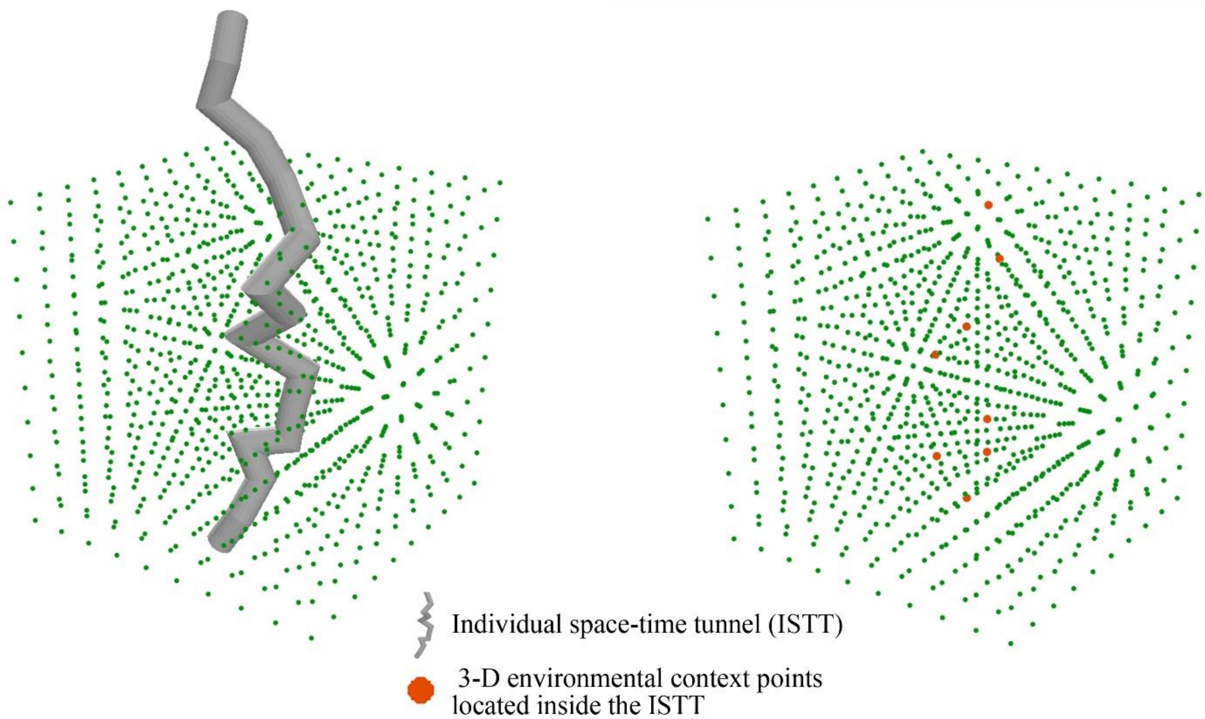


Figure 4.8 The 3-D intersection of the point cloud and individual space-time tunnel.

By identifying the 3-D intersection of the ISTT and the point cloud, subjects' exposures to the food environment can then be derived. After abstracting all intersected 3-D points from the ECC, the ECEI was evaluated as follows:

$$ECEI(j)(k) = \sum_{i=1}^{n} \frac{EC_{ij}W_i}{T} \quad (1 \leq i \leq n) \tag{4.1}$$

$$W_i = \begin{cases} 1 & if\ v_i = 0 \\ \left(\frac{1}{2}\right)^{v_i} & if\ v_i > 0 \end{cases} \quad (1 \leq i \leq \mathrm{n}) \tag{4.2}$$

where $ECEI(j)(k)$ is the environmental context exposure index of environmental factor j for subject k, EC_{ij} is the value of environmental factor j for 3-D point i based on the intersection of the ISTT and the ECC, W_i is the weight for 3-D point i, T is the time span of the intersection (the time unit can be hour or day), and n is the total number of points derived by the 3-D intersection.

W_i is a user-defined parameter for calculating the environmental context exposure index (ECEI). In most cases, the point weight (W_i) can be set to 1, but it can also differ for different research questions. One possible method for assigning voxel weight is to use movement velocity, as shown in Equation (2), where v_i is the movement velocity of the subject when passing through voxel i: The higher the velocity, the smaller the weight should be, indicating less contextual influences. When speed equals 0, the subject is staying at that location, so weight is set to 1. A quickly moving subject, by contrast, may pass by a location quite rapidly; thus, the influence of the environmental factor at 3-D point i should be small and the weight less than 1. Again, for purpose of illustration, W_i was set to 1 globally for this exploratory study.

Using this method, we calculated an environmental context exposure index (ECEI) for each of the 54 environmental context cubes (ECCs) separately. Because the three distance-decay methods generate ECCs with different value ranges, we standardized the ECEI by using z-score to facilitate the comparison of the exposures measured by different ECCs.

4.2.6 Comparing the Individual Food Environment Exposure Measurement with Other Methods

To compare the proposed analytical framework for food environment exposure measurement with other methods, four widely used exposure assessment methods (Shannon and Spurlock 1976, Arcury *et al.* 2005, Rainham *et al.* 2010, Zenk *et al.* 2011, Crawford *et al.* 2014, Kwan *et al.* 2018, Zhao *et al.* 2018) were also implemented with the same dataset. The four methods are GPS trajectory buffers (GTBs), standard deviation ellipses with one or two standard deviation(s) (SDE1, SDE2) and minimum convex polygons (MCPs). Figure 4.9 illustrates these four methods based on one subject's GPS trajectory (note that the geographic coordinates of the

GPS tracks shown in this figure have been modified for the purpose of human subjects' protection). The GTBs were created by a 100-meter 2-D buffer along the participant's GPS trajectories. The buffering area covered all the daily activity locations that this subject visited in the study period. The SDE is another widely-used method for exposure space delineation. Based on the transformed mean center and the rotated major and minor axes of all the GPS points of the subject, an ellipse was obtained based on either one or two standard deviation(s) of the distances between all pairs of GPS points. The SDEs represent the spatial distribution and directional trends of the subject's activity locations and normally does not include all of the GPS points. The MCP is the smallest convex polygon that contains all the GPS points of the subject, which covers all the daily activity locations. With the same data of BMI-unhealthy food outlets, exposure to BMI-unhealthy food environment was calculated as the density of BMI-unhealthy food outlets in these four delineations of exposure space. The results were standardized with z-score transformation and compared with those obtained using the ECCs and ECEIs.

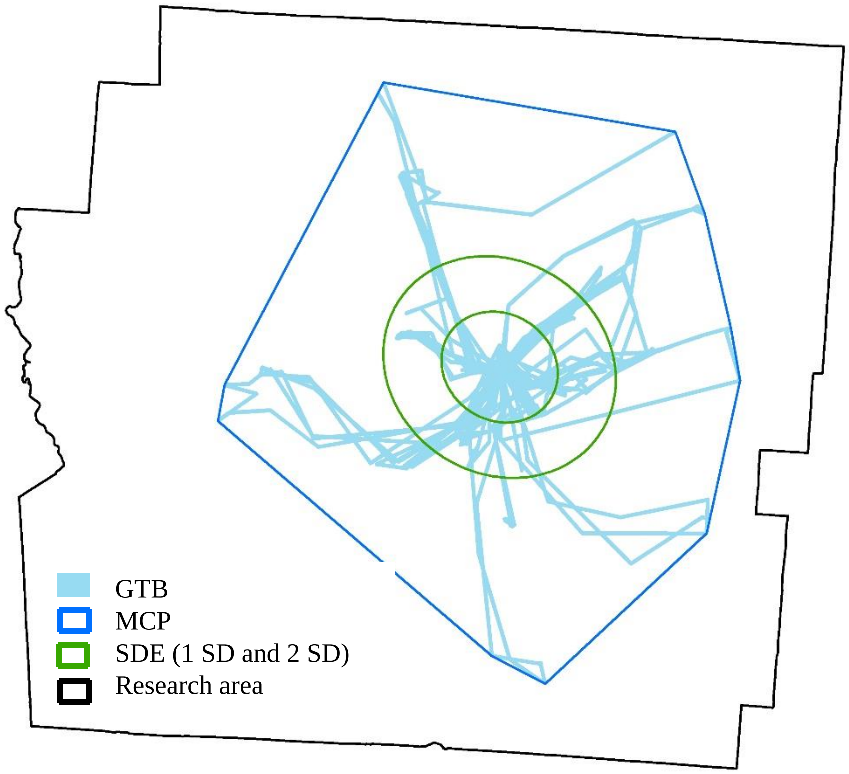


Figure 4.9 Four widely used methods for delineating individual exposure space based on one subject's GPS trajectory data.

4.2.7 Analytical Approach

To compare the performance of the environmental context cubes (ECCs) with different distance-decay methods at various spatial and temporal resolutions, as well as the exposure assessment results between the environmental context exposure index (ECEI) and other widely used food environment exposure measurements, we examined all these measurements and their relationship with participants' overweight status using binary logistic regression models, which are widely used in public health studies (Li *et al.* 2009, Zick *et al.* 2009, Mellor *et al.* 2011). The response variable is individual overweight status (0: non-overweight; 1: overweight or obese) based on participants' BMI (non-overweight: BMI < 25.0kg/m^2; overweight or obese: BMI ≥ 25.0 kg/m^2), while the independent variable is individual food environment exposure. The models were controlled for subjects' age, gender, and education level. A total of 22 models were built with different measurements of BMI-unhealthy food environment exposure (18 measured by the ECEI based on various ECCs and 4 measured by other methods). The performance of these models was compared by the Akaike information criterion (AIC), Nagelkerke R^2, likelihood ratio chi-square (LR χ^2) and the corresponding *p*-value, which indicate the robustness of the model. The more robust a model is, the better exposure assessment will be obtained. In addition, the models were compared to see if a significant association existed between individual food environment exposure and overweight status.

4.3 RESULTS

4.3.1 Variation in Food Environment Exposure Measurements with Different Methods

Individual food environment exposures were measured by the environmental context cubes (ECCs) and the environmental context exposure indexes (ECEIs) using three different distance-decay methods, three different spatial resolutions, and two different temporal resolutions, as well as other four widely used methods (GTB, MCP, SDE1, and SDE2). Figure 4.10 illustrates the standardized measures of these methods for each participant. In the figure, exposures measured by the ECCs with three distance-decay methods include only those with the highest spatial (100 m × 100 m) and temporal resolution (10 min), since they captured the finest detail of the spatial and temporal dynamics of the food environment. The measurement results of the ECCs with different spatial and temporal resolutions will be compared and discussed in the following sections. The horizontal axis of Figure 4.10 indicates the 46 participants in the study, while the vertical axis shows the food environment exposure measures. The figure indicates that different methods give considerably different exposure measures for the same participant.

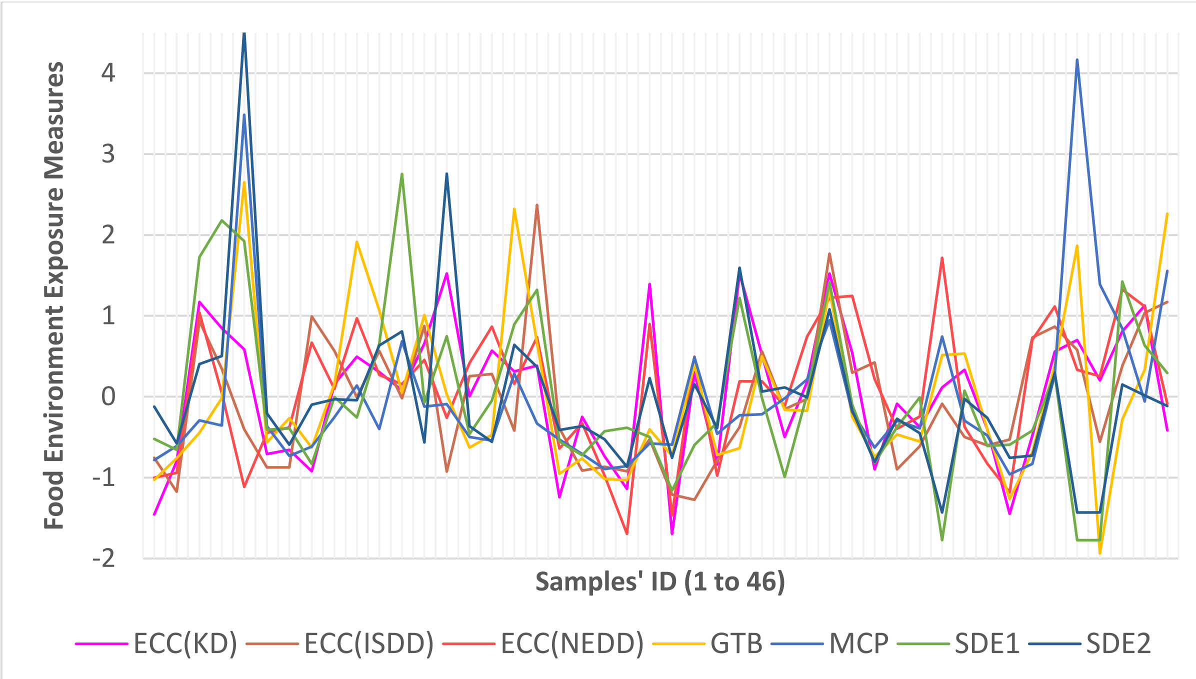


Figure 4.10 Comparison of the exposure measures obtained with different methods for each participant. (ECC: environmental context cube; KD: kernel density estimation; ISDD: inverse-square distance decay function; NEDD: negative-exponent distance decay function; GTB: GPS trajectory buffers; MCP: minimum convex polygons; SDE1: standard deviation ellipses with one standard deviation; SDE2: standard deviation ellipses with two standard deviations.)

To investigate the relationship among all the exposure measures obtained using different methods, we perform bivariate Pearson correlation analysis between each pair of the measures. Table 4.2 illustrates the results, which indicates that more than half of the pairs do not have significant correlations, including the pairs of ECC(KD) – MCP, ECC(ISDD) – MCP, ECC(ISDD) – SDE1, ECC(ISDD) – SDE2, ECC(NEDD) – GTB, ECC(NEDD) – MCP, ECC(NEDD) – SDE1, ECC(NEDD) – SDE2, GTB – SDE1, MCP – SDE1 and MCP – SDE2. Although the other pairs show significant correlations, most of the coefficients are smaller than 0.6, which indicate moderate to low associations. Only the pairs ECC(KD) – ECC(NEDD), ECC(ISDD) – ECC(NEDD), GTB – MCP and SDE1 – SDE2 shows strong associations. It is reasonable that the results of the ECCs with different distance-decay methods are correlated with each other since they share the same model and concepts, while SDE1 and SDE2 are also the same methods with various parameters. The results show that individual exposures measured by different methods are mostly different from and not correlated with each other, which indicates the existence of the UGCoP. Therefore, it is worth comparing these measurements and exploring the accurate ways to assess individual food environment exposures.

Table 4.2 The results of the bivariate Pearson correlation analysis between each pair of the methods for food environment exposure assessment.

Methods	ECC(KD)	ECC(ISDD)	ECC(NEDD)	GTB	MCP	SDE1	SDE2
ECC(KD)	-	0.407 *	0.658 *	0.438 *	0.364	0.508 *	0.476 *
ECC(ISDD)	-	-	0.642 *	0.396 *	0.198	0.358	0.057
ECC(NEDD)	-	-	-	0.269	0.180	0.168	-0.033
GTB	-	-	-	-	0.629 *	0.345	0.439 *
MCP	-	-	-	-	-	0.099	0.291
SDE1	-	-	-	-	-	-	0.699 *
SDE2	-	-	-	-	-	-	-

* *p*-value < 0.05

4.3.2 Comparing the Performance of Food Environment Exposure Measurement Methods

Table 4.3 shows the results of the binary logistic regression models with different measurements of individual BMI-unhealthy food environment exposure on the relationships between food environment exposure and overweight status. In the table, models KD100T10, KD100T30, KD150T10, KD150T30, KD200T10, and KD200T30 use exposures measured based on ECCs with KD as the distance-decay function in different spatial and temporal resolutions. In addition, models ISDD100T10, ISDD100T30, ISDD150T10, ISDD150T30, ISDD200T10, and ISDD200T30 use the exposure assessed based on ECCs with ISDD in various spatial and temporal resolutions. Furthermore, models NEDD100T10, NEDD100T30, NEDD150T10, NEDD150T30, NEDD200T10, and NEDD200T30 use exposure evaluated based on ECCs with NEDD in different spatial and temporal resolutions. Lastly, models M-GTB, M-MCP, M-SDE1, M-SDE2 use the measurement of individual food environment exposure based on four widely used methods (GTB, MCP, SDE1, and SDE2). The table shows that all the logistic regression models are statistically significant with p-value < 0.001. The models explained at least 45% (Nagelkerke R^2) of the variance of participants' overweight status. Among these models, the most robust one is the ISDD100T10 with the smallest AIC (45.552), largest LR χ^2 (28.217), and p-value < 0.001. The model explained 61.13% (Nagelkerke R^2) of the variance of participants' overweight status. The least robust model is the KD150T10 (AIC = 54.240, LR χ^2 = 19.529), which only explained 46.12% of the variance.

Table 4.3 The results of binary logistic regression models with different measurements of individual BMI-unhealthy food environment exposure for relationships between food environment exposure and overweight status.

Model [a, b]	Method	Spatial Resolution	Temporal Resolution	AIC	Nagelkerke R^2	LR χ^2	p-value
KD100T10	ECC (KD)	100 m × 100 m	10 min	53.898	0.4677	19.872	0.00053 ***
KD100T30			30 min	53.879	0.4681	19.891	0.00052 ***
KD150T10		150 m × 150 m	10 min	54.240	0.4612	19.529	0.00062 ***
KD150T30			30 min	54.228	0.4615	19.542	0.00061 ***
KD200T10		200 m × 200 m	10 min	54.198	0.4620	19.571	0.00061 ***
KD200T30			30 min	54.223	0.4616	19.546	0.00061 ***
ISDD100T10	ECC (ISDD)	100 m × 100 m	10 min	45.552	0.6113	28.217	0.00001 ***
ISDD100T30			30 min	48.567	0.5624	25.202	0.00005 ***
ISDD150T10		150 m × 150 m	10 min	53.272	0.4794	20.498	0.00040 ***
ISDD150T30			30 min	52.801	0.4881	20.969	0.00032 ***
ISDD200T10		200 m × 200 m	10 min	53.124	0.4822	20.645	0.00037 ***
ISDD200T30			30 min	54.073	0.4644	19.696	0.00057 ***
NEDD100T10	ECC (NEDD)	100 m × 100 m	10 min	49.769	0.5420	24.001	0.00008 ***
NEDD100T30			30 min	53.597	0.4734	20.172	0.00046 ***
NEDD150T10		150 m × 150 m	10 min	52.945	0.4855	20.825	0.00034 ***
NEDD150T30			30 min	51.792	0.5064	21.977	0.00020 ***
NEDD200T10		200 m × 200 m	10 min	54.132	0.4633	19.638	0.00059 ***
NEDD200T30			30 min	54.199	0.4650	19.571	0.00061 ***
M-GTB	GTB	-	-	51.462	0.5124	22.308	0.00017 ***
M-MCP	MCP	-	-	52.390	0.4956	21.380	0.00027 ***
M-SDE1	SDE1	-	-	53.542	0.4744	20.228	0.00045 ***
M-SDE2	SDE2	-	-	51.753	0.5071	22.016	0.00020 ***

[a] N = 46 subjects, [b] Response variable: 0 non-overweight; 1 overweight, *** p-value < 0.001, AIC: Akaike information criterion (the smaller the value, the better the model fit), LR χ^2: likelihood ratio chi-square (the larger the value, the better the model fit)

Among the ECC models, the food environment exposures estimated by ISDD generates the best results, while the ones estimated by KD generate the worst results. Considering various spatial and temporal resolutions of the ECC, the finer the resolution, the better the results. For instance, among the ECC models with specific distance-decay function, the most robust model is always the one with the finest spatial resolution: KD100T30 (AIC = 53.879, Nagelkerke R^2 = 0.4681, LR χ^2 = 19.891) for ECC(KD); ISDD100T10 (AIC = 45.552, Nagelkerke R^2 = 0.6113, LR

χ^2 = 28.217) for ECC(ISDD); NEDD100T10 (AIC = 49.769, Nagelkerke R^2 = 0.5420, LR χ^2 = 24.001) for ECC(NEDD). In addition, the table indicates that ECCs with a temporal resolution of 10 min normally generate better results compared to the ones with a temporal resolution of 30 min with several exceptions.

Regarding the four widely used exposure assessment methods, GTB performs the best with AIC = 51.462, Nagelkerke R^2 = 0.5124, LR χ^2 = 22.308. However, the models with ECC(ISDD) and ECC(NEDD) still perform much better than the GTB-based model. It is worth noting that the models with ECC(KD) have the worst performance when compared to all the other methods, which may indicate that the effects of food outlets may not follow the decay patterns as depicted by a kernel density estimation.

4.3.3 Association between Food Environment Exposure and Overweight Status

The associations between food environment exposure based on different ECCs and participants' overweight status are shown in Table 4.4. Almost all the models, except NEDD150T10, indicate that being female (compared to being male) is associated with higher odds of being overweight, while having a college degree and higher (compared high school degree and lower) is associated with lower odds of being overweight. However, a significant association between BMI-unhealthy food environment exposure and overweight status is observed for only three of the models (ISDD100T10, ISDD100T30, and NEDD100T10). Higher unhealthy food environment exposure is found to be significantly associated with higher odds of being overweight in models ISDD100T10 (odds ratio (OR): 6.81; 95% confidential interval (CI): 1.76, 45.3; p-value < 0.01), ISDD100T30 (OR: 4.35; CI: 1.27, 22.62; p-value < 0.1) and NEDD100T10 (OR: 3.13; CI: 1.08, 11.47; p-value < 0.1). Referring to the performance of models discussed above, these three models are also the most robust models with the lowest AIC and highest LR χ^2, as well as the highest explanation rate of the variance in participants' overweight status.

Table 4.4 The association between food environment exposure and overweight status analyzed by binary logistic regression models with food environment exposure measured by different ECCs.

Variables	Model [a, b]	β (95% CI)	OR (95% CI)	Model [a, b]	β (95% CI)	OR (95%)
Gender (Female)	KD100T10	2.56 *** (0.97, 4.63)	12.97 *** (2.64, 102.57)	KD100T30	2.57 ** (0.97, 4.63)	13.01 ** (2.65, 102.77)
Age		0.05 (−0.02, 0.14)	1.05 (0.98, 1.14)		0.05 (−0.02, 0.14)	1.05 (0.98, 1.15)
Education (≥ College Degree)		−2.37 ** (−4.46, −0.72)	0.09 ** (0.01, 0.49)		−2.37 ** (−4.47, −0.73)	0.09 ** (0.01, 0.48)
Env. Exp.		0.28 (−0.64, 1.27)	0.09 (0.01, 0.49)		0.29 (−0.64, 1.28)	1.34 (0.53, 3.60)
Gender (Female)	KD150T10	2.53 *** (0.91, 4.63)	12.62 *** (2.49, 102.72)	KD150T30	2.54 ** (0.92, 4.64)	12.73 ** (2.51, 103.68)
Age		0.05 (−0.02, 0.14)	1.05 (0.98, 1.15)		0.05 (−0.02, 0.14)	1.05 (0.98, 1.14)
Education (≥ College Degree)		−2.43 *** (−4.56, −0.76)	0.09 *** (0.01, 0.47)		−2.42 ** (−4.55, −0.76)	0.09 ** (0.01, 0.47)
Env. Exp.		0.06 (−0.94, 1.04)	1.06 (0.39, 2.82)		0.08 (−0.90, 1.04)	1.08 (0.41, 2.83)
Gender (Female)	KD200T10	2.54 *** (0.94 4.62)	12.68 *** (2.56, 101.55)	KD200T30	2.53 ** (0.94, 4.61	12.57 ** (2.56, 100.23)
Age		0.05 (−0.02, 0.13)	1.05 (0.98, 1.14)		0.05 (−0.02, 0.14)	1.05 (0.98, 1.14)
Education (≥ College Degree)		−2.44 *** (−4.53, −0.83)	0.09 *** (0.01, 0.44)		−2.45 ** (−4.53, −0.83)	0.09 ** (0.01, 0.43)
Env. Exp.		0.11 (−0.84, 1.02)	1.12 (0.43, 2.77)		0.08 (−0.87, 0.99)	1.09 (0.42, 2.69)
Gender (Female)	ISDD100T10	3.61 *** (1.58, 6.37)	36.83 *** (4.87, 584.49)	ISDD100T30	3.30 ** (1.41, 5.84)	27.13 ** (4.10, 342.79)
Age		0.03 (−0.04, 0.12)	1.04 (0.96, 1.13)		0.03 (−0.05, 0.12)	1.03 (0.95, 1.12)
Education (≥ College Degree)		−2.13 ** (−4.42, −0.31)	0.12 ** (0.01, 0.74)		−1.93 * (−4.09, −0.19)	0.14 * (0.02, 0.83)
Env. Exp.		1.92 ** (0.57, 3.81)	6.81 ** (1.76, 45.3)		1.47 * (0.24, 3.12)	4.35 * (1.27, 22.62)
Gender (Female)	ISDD150T10	2.90 *** (1.12, 5.23)	18.14 *** (3.06, 186.21)	ISDD150T30	2.92 ** (1.17, 5.21)	18.48 ** (3.21, 182.24)
Age		0.04 (−0.03, 0.13)	1.04 (0.97, 1.14)		0.04 (−0.03, 0.13)	1.05 (0.97, 1.14)
Education (≥ College Degree)		−2.34 ** (−4.42, −0.71)	0.10 ** (0.01, 0.49)		−2.36 * (−4.45, −0.72)	0.09 * (0.01, 0.48)
Env. Exp.		0.45 (−0.44, 1.42)	1.57 (0.64, 4.12)		0.51 (−0.31, 1.47)	1.67 (0.73, 4.36)
Gender (Female)	ISDD200T10	2.68 *** (1.03, 4.86)	14.61 *** (2.80, 129.08)	ISDD200T30	2.57** (0.96, 4.68)	13.09 ** (2.62, 107.41)
Age		0.05 (−0.02, 0.13)	1.05 (0.98, 1.14)		0.05 (−0.02, 0.13)	1.05 (0.98, 1.14)
Education (≥ College Degree)		−2.60 *** (−4.77, −0.94)	0.07 *** (0.01, 0.39)		−2.47 ** (−4.55, −0.85)	0.09 ** (0.01, 0.43)
Env. Exp.		0.49 (−0.41, 1.46)	1.63 (0.66, 4.31)		0.19 (−0.69, 1.11)	1.21 (0.50, 3.04)
Gender (Female)	NEDD100T10	2.81 *** (1.10, 5.04)	16.55 *** (3.02, 154.12)	NEDD100T30	2.57 ** (0.98, 4.64)	13.07 ** (2.66, 103.57)
Age		0.05 (−0.03, 0.13)	1.05 (0.97, 1.14)		0.05 (−0.02, 0.13)	1.05 (0.98, 1.14)
Education (≥ College Degree)		−1.98 ** (−4.13, −0.26)	0.14 ** (0.02, 0.77)		−2.20 * (−4.35, +0.49)	0.11 * (0.01, 0.61)
Env. Exp.		1.14 * (0.08, 2.44)	3.13 * (1.08, 11.47)		0.37 (−0.51, 1.40)	1.44 (0.60, 4.06)

Table 4.4 (continued)

Variables	Model [a, b]	β (95% CI)	OR (95% CI)	Model [a, b]	β (95% CI)	OR (95%)
Gender (Female)	NEDD150T10	2.90 *** (1.15, 5.19)	18.25 *** (3.17, 180.20)	NEDD150T30	3.00 ** (1.25, 5.29)	20.09 ** (3.49, 199.09)
Age		0.04 (−0.04, 0.12)	1.04 (0.96, 1.13)		0.04 (−0.04, 0.12)	1.04 (0.96, 1.13)
Education (≥ College Degree)		−2.19 (−2.29, −0.52)	0.11 (0.01, 0.59)		−2.21 * (−4.31, −0.55)	0.11 * (0.01, 0.58)
Env. Exp.		0.57 (−0.40, 1.62)	1.76 (0.67, 5.06)		0.72 (−0.17, 1.77)	2.06 (0.84, 5.81)
Gender (Female)	NEDD200T10	2.53 *** (0.94, 4.62)	12.64 *** (2.57, 101.77)	NEDD200T30	2.50 ** (0.92, 4.57)	12.22 ** (2.51, 96.49)
Age		0.05 (−0.02, 0.13)	1.05 (0.98, 1.14)		0.05 (−0.02, 0.14)	1.06 (0.98, 1.15)
Education (≥ College Degree)		−2.46 *** (−4.54, −0.85)	0.09 *** (0.01, 0.43)		−2.48 ** (−4.57, −0.86)	0.08 ** (0.01, 0.42)
Env. Exp.		0.14 (−0.68, 0.93)	1.15 (0.51, 2.54)		−0.09 (−0.89, 0.67)	0.91 (0.41, 1.95)

[a] N = 46 subjects, [b] Response variable: 0 non-overweight; 1 overweight, * *p*-value < 0.1, ** *p*-value < 0.01, *** *p*-value < 0.001, Gender: 0-male, 1-female, Education: 0-without a college degree or higher; 1-with a college degree or higher, Env. Exp.: BMI-unhealthy food environment exposure measured by a specific method.

Table 4.5 lists the results of the binary logistic regression models based on the other four widely used methods. The models M-GTP, M-MCP and M-SDE1 indicate that being female is associated with higher odds of being overweight while having a college degree and higher is associated with lower odds of being overweight. M-SDE2 is the only model that does not find an association between education level and overweight status. Interestingly, all these four models did not find any significant association between BMI-unhealthy food environment exposure and participants' overweight status.

Table 4.5 The association between food environment exposure and overweight status analyzed by the binary logistic regression models with the food environment exposure measured by the other four widely used methods.

Variables.	Model [a, b]	β (95% CI)	OR (95% CI)
Gender (Female)	M-GTP	3.12 *** (1.29, 5.62)	22.68 *** (3.62, 275.66)
Age	M-GTP	0.08 (0.00, 0.17)	1.08 (10.00, 1.19)
Education (≥ College Degree)	M-GTP	−2.48 ** (−4.72, −0.76)	0.08 ** (8.91, 0.47)
Env. Exp.	M-GTP	0.29 (−0.05, 0.69)	1.34 (9.52, 2.00)
Gender (Female)	M-MCP	2.57 *** (0.95, 4.70)	13.10 *** (2.60, 109.75)
Age	M-MCP	0.06 (−0.01, 0.15)	1.06 (0.99, 1.17)
Education (≥ College Degree)	M-MCP	−2.53 *** (−4.70, −0.86)	0.08 *** (0.01, 0.42)
Env. Exp.	M-MCP	0.56 (−0.23, 1.53)	1.74 (0.79, 4.60)
Gender (Female)	M-SDE1	2.32 ** (0.66, 4.43)	10.16 ** (1.93, 83.83)
Age	M-SDE1	0.06 (−0.02, 0.14)	1.06 (0.98, 1.15)
Education (≥ College Degree)	M-SDE1	−2.74 *** (−5.03, −0.99)	0.06 *** (0.01, 0.37)
Env. Exp.	M-SDE1	−0.24 (−0.90, 0.31)	0.78 (0.40, 1.36)
Gender (Female)	M-SDE2	2.33 ** (0.64, 4.51)	10.26 ** (1.90, 90.62)
Age	M-SDE2	0.04 (−0.03, 0.13)	1.05 (0.97, 1.14)
Education (≥ College Degree)	M-SDE2	−2.83 (−5.17, −1.07)	0.06 (0.01, 0.34)
Env. Exp.	M-SDE2	−0.57 (−1.36, 0.12)	0.57 (0.26, 1.12)

[a] N = 46 subjects, [b] Response variable: 0 non-overweight; 1 overweight, ** *p*-value < 0.01, *** *p*-value < 0.001.

The results indicate that the proposed framework generates better measurements of individual food environment exposures when compared to other widely used methods. This suggests the inconsistent findings in previous studies may be partly due to the methods used. Significant associations between BMI-unhealthy food environment exposures and overweight status were found in the three most robust models (ISDD100T10, ISDD100T30, and NEDD100T10). Being the most robust model, ISDD100T10 (explained 61.13% of the variance in participants' overweight status) found that higher unhealthy food environment exposure (OR: 6.81; CI: 1.76, 45.3; *p*-value < 0.01) is significantly associated with higher odds of being overweight.

4.4 DISCUSSION

The proposed framework generated more reasonable and reliable results when compared to other methods, and thus obtained more accurate individual food environment exposures assessment. Regarding the distance-decay methods for generating the ECC, the ISDD represents the dynamic environmental contexts more accurately, and the ECC(ISDD) with a spatial resolution of 100 m × 100 m and a temporal resolution of 10 min performs best with the most robust regression models. Regarding the spatial and temporal resolution of the ECC, the finer the spatial and temporal resolution, the better the performance of the model. This suggests the existence of scale effects when using the ECC for measuring individual exposures. Thus, future application of the ECC may need to consider a proper spatial and temporal resolution in order to generate reliable results.

The framework proposed in this study can help to mitigate the UGCoP. With respect to the spatial dimension, contextual units or areas in the study were not based on arbitrary predefined spatial boundaries (e.g., census tracts) but were delineated by ISTTs based on participants' actual movement trajectories (GPS tracks). This is significantly different from the methods used in most previous studies, which tended to measure contextual influences based on static residential neighborhoods that may not accurately represent the actual areas that exert contextual influences on individual behavior or health outcomes (Matthews 2008, Chaix 2009, Kwan 2009a, 2018a). On the other hand, temporal variations in relevant environmental contexts were handled dynamically: The influence of the food outlets was measured with consideration of their business hours in order to more accurately capture their contextual effects. Taking into account the complex spatial and temporal configuration of individual contextual exposure, the proposed framework and methods help to mitigate UGCoP.

The ECEI based on the ECC and ISTT in the study is a quantitative measure of individual exposure to environmental contexts in unit time, which may be used as a standard measure of individual contextual exposure. It provides a useful standardized tool for environmental exposure assessment, while capturing the spatial and temporal variations in environmental contexts. It is also flexible to implement the ECEI for different research questions concerning different environmental contexts using the two user-defined parameters B_r and W_i. These two parameters can be explored in further studies to fit the research question. The index may be further used for comparative analysis of environmental exposures between different individuals or groups. In addition, the ECEI may be utilized to examine the relationships between environmental contexts and other health outcomes. The observed association may be used to investigate and identify high-risk environmental contexts and provide decision support for policy-making in public health.

Lastly, this research has several limitations that need to be addressed in future studies. First, this study implemented the proposed framework based on a relatively small sample of participants who live near parks. Larger GPS datasets with more subjects from different study sites are thus needed in future research to further evaluate the robustness of the framework. Second, this study only applied the framework to food environment exposures; further studies are needed to assess its effectiveness for addressing other health issues, such as physical activity and mental health. Third, we implemented the three distance-decay methods in Euclidean distance without considering transport modes and the configuration of road networks. More sophisticated methods (Apparicio *et al.* 2017) that incorporate transport modes in the ECC would help to further the application of the framework and has significant potential to better represent the environmental context. We will develop the ECC along this line in future studies. Fourth, since there is no data on participants' actual activities, there may be some uncertainty in the exposure measure. For instance, working and eating at a fast-food restaurant may mean different exposure and have different effects on a participant’s body weight. If activity diary data are available, activity types can be integrated into the calculation of the ECEI by differentiating the contextual effects of different types of activity. Fifth, the proposed methodology can only explore the association between environmental exposures and health outcomes. Further investigations (e.g., controlled experiments or longitudinal studies) are still needed to validate any causal relationships.

CHAPTER 5: CONCLUSION AND FUTURE DIRECTIONS

5.1 SUMMARY OF CONTRIBUTIONS

This dissertation explores methodological issues in environmental health research that cause uncertain findings, with a particular focus on how the UGCoP influences research findings. As it does, it proposes activity space–based approaches to comprehensively assessing individual exposure to environmental contexts and proposes an innovative environmental exposure evaluation framework for spatiotemporally assessing individual environmental exposure and evaluating environmental effects on health outcomes. Based on empirical analysis of real-world applications using GPS tracking data and environmental context data collected in Chicago, Illinois, and Columbus, Ohio, these proposed approaches outperform other methods currently in wide use and mitigate the effects of the UGCoP.

The first piece of work (chapter 2) proposed a hexagon-based adaptive crystal growth Voronoi diagram–based approach that extends the adaptive crystal growth Voronoi diagrams by using a hexagon-based spatial framework instead of a raster grid structure to represent environmental context and demand for service. In this study, use of a different representation of space (i.e., continuous space covered with hexagonal cells) led to an improvement of service area delineation results. The results of the middle school service area delineation reveal the advantage of hexagon-based adaptive crystal growth Voronoi diagrams over the raster-based method, considering both the degree to which the population in each service area is commensurate with the enrollment capacity of the corresponding middle school and the degree to which the middle school is accessible within its service area. In the hexagon-based method, the estimated continuous socioeconomic weighted plane mitigates the MAUP, and the hexagon grid accessibility-weighted plane suffers less from orientation bias and sampling bias caused by edge effects, potentially helping delimit service areas that better match the service capacities of public facilities and that improve the accessibility of public facilities within their service areas. Future case studies of different kinds of public facilities in various settings and contexts could shed light on the generalizability of the conclusion that HACG works better than RACG for service area delineation. The proposed method could provide decision support for policymakers in city management and urban planning (e.g., educational administration) based on robust quantitative analysis results. Although the hexagon-based adaptive crystal growth Voronoi diagrams proposed in this study were implemented for the allocation of public service resources, they can also be used in environmental health research to delineate individual activity space based on subjects' daily movement patterns as well as environmental context, represented as hexagon-grid weighted planes.

Based on the proposed hexagon-based adaptive crystal growth Voronoi diagrams, the second piece of work (chapter 3) develops an innovative method for delineating activity space using individual GPS trajectories and a crystal-growth algorithm based on hexagon-grid accessibility-weighted planes. Such an approach generates a more reasonable activity space and captures individual environmental exposures more accurately than other methods do. This new tool for activity space delineation, which can be used to explore the relationships between human movement patterns and environmental context as well as environmental effects on health outcomes, is potentially groundbreaking in its introduction and development of a new analytical framework allowing the examination of activity space and individual environmental exposures while mitigating the UGCoP.

The third piece of work (chapter 4) establishes an analytical framework for dynamically representing food environment and deriving individual environmental exposures that effectively integrate human movement in space and time (e.g., GPS trajectories). Though designed to examine individual food environment exposure, this proposed framework, which incorporates the dynamics of a food environment into the ECC, captures individual exposure space using the individual space–time tunnel (ISTT), and assesses the effects of individual exposure on people's overweight status using the environmental context exposure index (ECEI), can also be used in a wide range of environmental health studies.

In summary, the contributions of this dissertation are both practical and methodological in nature. This research is of practical interest to urban planners, city policymakers, and community health practitioners who seek to stem the rise of environmental exposure–related chronic diseases, such as obesity and cancer, and to promote public health. The ability to more accurately profile environmental context and assess environmental exposures could help researchers investigate the association between health outcomes and environmental context with a view to developing policies and interventions that promote a healthful environment. Beyond environmental interventions and policies, development of innovative exposure and accessibility assessment methods could inform new avenues for individually tailored interventions, such as real-time warnings of unhealthy exposures via mobile devices.

Methodologically, this dissertation makes several contributions to environmental health studies, offering new tools for spatiotemporally profiling environmental contexts as well as innovative methods for assessing environmental exposures—accurately profiled contexts and assessments of environmental exposure being two key prerequisites for environmental effects analysis. In addition, its conclusions could help scholars analyze and understand the methodological uncertainty associated with the inconsistent findings of previous research as a result of the UGCoP and the spatial non-stationarity problem. By improving understandings of environmental effects on different health behaviors and outcomes, this research thus provides a solid benchmark for future research.

5.2 FUTURE STUDIES

In its current form, this dissertation could be expanded in several interrelated directions through subsequent studies. First, whereas this study implements the proposed methods based on the GPS trajectories of a small sample of participants collected during a relatively short period in Chicago, Illinois, and Columbus, Ohio, a larger dataset featuring the GPS trajectories of more subjects from different study sites over a longer tracking period would allow further evaluation of the robustness of the proposed approaches and provide justification for their use. In addition, because this dissertation investigated environmental exposure and its relationship to subjects' levels of physical activity and overweight status, further application of the proposed technique for environmental exposure assessment to various environmental factors and health outcomes is needed.

Second, although the methods presented in this study were compared with other existing approaches also based on GPS tracking data, still other methods do not rely on GPS or GIS data. For example, studies that have used map-based electronic questionnaires, mobility surveys, and activity space questionnaires have also yielded useful results. Although these methods are based on self-reported information and may thus incorporate recall bias and not be geographically accurate, they can capture background information about participants' activities, such as that relating to transportation modes and social interactions. Accordingly, further studies should compare or integrate these methods with the proposed CCG activity space and ECC analysis framework with a view to generating more accurate exposure assessment.

Third, the proposed analytical framework implemented the three distance decay methods in Euclidean distance without considering transport modes and road network configuration. More sophisticated methods (Apparicio *et al.* 2017) incorporating transport modes into the ECC would advance applications of the framework and would have significant potential to better represent environmental context. Furthermore, the ECC used the cube as the basic voxel for representing environmental context at a specific time and location, but use of a hexagonal prism instead might better represent the environmental context, capitalizing on the benefits of the hexagonal grid in spatial analysis. We intend to develop and improve the ECC along these lines in future studies.

Finally, because the proposed methods did not consider participants' actual activities at different locations, some uncertainty may be attached to the exposure measure. For example, working and eating at a fast-food restaurant might well be associated with exposures of different natures that could differ in their effects on body weight. In future studies, should activity diary data be available, activity types could be integrated into calculations of individual environmental exposures by differentiating the contextual effects of different activity types. Furthermore, the proposed methods are able to explore only the association between environmental exposures and health outcomes; further investigations (e.g., controlled experiments or longitudinal studies) would be needed to validate any causal relationships.

REFERENCES

aan het Rot, M., Hogenelst, K., and Schoevers, R.A., 2012. Mood disorders in everyday life: a systematic review of experience sampling and ecological momentary assessment studies. *Clinical psychology review*, 32 (6), 510–523.

Acevedo-Garcia, D., 2000. Residential segregation and the epidemiology of infectious diseases. *Social science & medicine*, 51 (8), 1143–1161.

Adams, B. and Kapan, D.D., 2009. Man bites mosquito: Understanding the contribution of human movement to vector-borne disease dynamics. *PLoS ONE*, 4 (8), 6763.

Adams, M.A., Sallis, J.F., Kerr, J., Conway, T.L., Saelens, B.E., Frank, L.D., Norman, G.J., and Cain, K.L., 2011. Neighborhood environment profiles related to physical activity and weight status: a latent profile analysis. *Preventive medicine*, 52 (5), 326–331.

Aerts, J.C.J.H. and Heuvelink, G.B.M., 2002. Using simulated annealing for resource allocation. *International Journal of Geographical Information Science*, 16 (6), 571–587.

Aichholzer, O., Aurenhammer, F., and Palop, B., 2004. Quickest Paths, Straight Skeletons, and the City Voronoi Diagram. *Discrete and Computational Geometry*, 31 (1), 17–35.

Albert, D.P. and Gesler, W.M., 2003. How spatial analysis can be used in medical geography. *In*: *Spatial Analysis, GIS and Remote Sensing*. CRC Press, 19–46.

Almanza, E., Jerrett, M., Dunton, G., Seto, E., and Ann Pentz, M., 2012. A study of community design, greenness, and physical activity in children using satellite, GPS and accelerometer data. *Health and Place*, 18 (1), 46–54.

Aly, A.A. and Marucheck, A.S., 1982. Generalized Weber problem with rectangular regions. *Journal of the Operational Research Society*, 33 (11), 983–989.

Andersen, A.F., Carson, C., Watt, H.C., Lawlor, D.A., Avlund, K., and Ebrahim, S., 2008. Life-course socio-economic position, area deprivation and type 2 diabetes: Findings from the British women's heart and health study. *Diabetic Medicine*, 25 (12), 1462–1468.

Aneshensel, C.S. and Sucoff, C.A., 1996. The Neighborhood Context of Adolescent Mental Health. *Journal of Health and Social Behavior*, 37 (4), 293–310.

Anyamba, A., Linthicum, K.J., Mahoney, R., Tucker, C.J., and Kelley, P.W., 2002. Mapping potential risk of Rift Valley fever outbreaks in African savannas using vegetation index time series data. *Photogrammetric engineering and remote sensing*, 68 (2), 137–145.

Apparicio, P., Gelb, J., Dubé, A.S., Kingham, S., Gauvin, L., and Robitaille, É., 2017. The approaches to measuring the potential spatial access to urban health services revisited: Distance types and aggregation-error issues. *International Journal of Health Geographics*, 16 (32), 1–24.

Arcury, T.A., Gesler, W.M., Preisser, J.S., Sherman, J., Spencer, J., and Perin, J., 2005. The effects of geography and spatial behavior on health care utilization among the residents of a rual region. *Health Services Research*, 40 (1), 135–155.

Arem, H., Moore, S.C., Patel, A., Hartge, P., Berrington De Gonzalez, A., Visvanathan, K., Campbell, P.T., Freedman, M., Weiderpass, E., Adami, H.O., Linet, M.S., Lee, I.M., and Matthews, C.E., 2015. Leisure time physical activity and mortality: A detailed pooled analysis of the dose-response relationship. *JAMA Internal Medicine*, 175 (6), 959–967.

Auchincloss, A.H., 2009. Neighborhood Resources for Physical Activity and Healthy Foods and Incidence of Type 2 Diabetes Mellitus. *Archives of Internal Medicine*, 169 (18), 1698.

Basta, L. a., Richmond, T.S., and Wiebe, D.J., 2010. Neighborhoods, daily activities, and measuring health risks experienced in urban environments. *Social Science and Medicine*, 71 (11), 1943–1950.

Bavia, M.E., Malone, J.B., Hale, L., Dantas, A., Marroni, L., and Reis, R., 2001. Use of thermal and vegetation index data from earth observing satellites to evaluate the risk of schistosomiasis in Bahia, Brazil. *Acta tropica*, 79 (1), 79–85.

Benarous, X., Edel, Y., Consoli, A., Brunelle, J., Etter, J.-F., Cohen, D., and Khazaal, Y., 2016. Ecological momentary assessment and smartphone application intervention in adolescents with substance use and comorbid severe psychiatric disorders: study protocol. *Frontiers in psychiatry*, 7, 157.

Bennett, C.D. and Mirakhor, A., 1974. Optimal facility location with respect to several regions. *Journal of Regional Science*, 14 (1), 131–136.

Benzarti, E., Sahin, E., and Dallery, Y., n.d. Modeling Approaches for the home health care districting problem. *In*: *8th International Conference of Modeling and Simulation-MOSIM*. Hammamet.

Berke, E.M., Koepsell, T.D., Moudon, A. V, Hoskins, R.E., and Larson, E.B., 2007. Association of the built environment with physical activity and obesity in older persons. *American journal of public health*, 97 (3), 486–492.

Birch, C.P.D., Oom, S.P., and Beecham, J.A., 2007. Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. *Ecological Modelling*, 206 (3–4), 347–359.

Black, J.L., Macinko, J., Dixon, L.B., and Fryer, G.E., 2010. Neighborhoods and obesity in New York City. *Health and Place*, 16 (3), 489–499.

Boardman, J.D., Finch, B.K., Ellison, C.G., Williams, D.R., and Jackson, J.S., 2001. Neighborhood disadvantage, stress, and drug use among adults. *Journal of health and social behavior*, 42 (2), 151–165.

Boots, B. and South, R., 1997. Modeling retail trade areas using higher-order, multiplicatively weighted voronoi diagrams. *Journal of Retailing*, 73 (4), 519–536.

Boyle, P.J. and Dunn, C.E., 1991. Redefinition of Enumeration District Centroids: A Test of Their Accuracy by Using Thiessen Polygons. *Environment and Planning A*, 23 (8), 1111–1119.

Bozkaya, B., Erkut, E., and Laporte, G., 2003. A tabu search heuristic and adaptive memory procedure for political districting. *European Journal of Operational Research*, 144 (1), 12–26.

Braza, M., Shoemaker, W., and Seeley, A., 2004. Neighborhood design and rates of walking and biking to elementary school in 34 California communities. *American journal of health promotion*, 19 (2), 128–136.

Brown, D.W., Balluz, L.S., Heath, G.W., Moriarty, D.G., Ford, E.S., Giles, W.H., and Mokdad, A.H., 2003. Associations between recommended levels of physical activity and health-related quality of life Findings from the 2001 Behavioral Risk Factor Surveillance System (BRFSS) survey. *Preventive medicine*, 37 (5), 520–528.

Browning, M. and Lee, K., 2017. Within what distance does "Greenness" best predict physical health? A systematic review of articles with GIS buffer analyses across the lifespan. *International journal of environmental research and public health*, 14 (7), 675.

Brunsdon, C., Fotheringham, A.S., and Charlton, M.E., 1996. Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical analysis*, 28 (4), 281–298.

Buck, C., Kneib, T., Tkaczick, T., Konstabel, K., and Pigeot, I., 2015. Assessing opportunities for physical activity in the built environment of children: Interrelation between kernel density and neighborhood scale. *International Journal of Health Geographics*, 14 (1), 1–16.

Caro, F., Shirabe, T., Guignard, M., and Weintraub, A., 2004. School Redistricting: Embedding GIS Tools with Integer Programming. *The Journal of the Operational Research Society*, 55 (8), 836–849.

Carr, D.B., Olsen, A.R., and White, D., 1992. Hexagon Mosaic Maps for Display of Univariate and Bivariate Geographical Data. *Cartography and Geographic Information Science*, 19 (4), 228–236.

Carver, A., Timperio, A., and Crawford, D., 2008a. Playing it safe: The influence of neighbourhood safety on children's physical activity-A review. *Health & place*, 14 (2), 217–227.

Carver, A., Timperio, A., and Crawford, D., 2008b. Neighborhood road environments and physical activity among youth: the CLAN study. *Journal of Urban Health*, 85 (4), 532.

Caspi, C.E., Sorensen, G., Subramanian, S. V., and Kawachi, I., 2012. The local food environment and diet: A systematic review. *Health and Place*, 18 (5), 1172–1187.

Cattaneo, A., Taronna, M., Garramone, G., Peruzzo, C., Schlitt, C., Consonni, D., and Cavallo, D.M., 2010. Comparison between personal and individual exposure to Urban air pollutants. *Aerosol Science and Technology*, 44 (5), 370–379.

Chaix, B., 2009. Geographic life environments and coronary heart disease: a literature review, theoretical contributions, methodological updates, and a research agenda. *Annual review of public health*, 30, 81–105.

Chaix, B., Kestens, Y., Duncan, D.T., Brondeel, R., Méline, J., El Aarbaoui, T., Pannier, B., and Merlo, J., 2016. A GPS-Based Methodology to Analyze Environment-Health Associations at the Trip Level: Case-Crossover Analyses of Built Environments and Walking. *American*

Journal of Epidemiology, 184 (8), 579–589.

Chaix, B., Kestens, Y., Perchoux, C., Karusisi, N., Merlo, J., and Labadi, K., 2012. An Interactive Mapping Tool to Assess Individual Mobility Patterns in Neighborhood Studies. *American Journal of Preventive Medicine,* 43 (4), 440–450.

Chaix, B., Méline, J., Duncan, S., Jardinier, L., Perchoux, C., Vallée, J., Merrien, C., Karusisi, N., Lewin, A., Brondeel, R., and Kestens, Y., 2013. Neighborhood environments, mobility, and health: Towards a new generation of studies in environmental health research. *Revue d'Epidemiologie et de Sante Publique*, 61 (SUPPL.3), 139–145.

Chaix, B., Méline, J., Duncan, S., Merrien, C., Karusisi, N., Perchoux, C., Lewin, A., Labadi, K., and Kestens, Y., 2013. GPS tracking in neighborhood and health studies: A step forward for environmental exposure assessment, A step backward for causal inference? *Health and Place*, 21, 46–51.

Chaix, B., Merlo, J., Evans, D., Leal, C., and Havard, S., 2009. Neighbourhoods in eco-epidemiologic research: delimiting personal exposure areas. A response to Riva, Gauvin, Apparicio and Brodeur. *Social science & medicine*, 69 (9), 1306–1310.

Chen, A.Y., Yu, T.-Y., Lu, T.-Y., Chuang, W.-L., Lai, J.-S., Yeh, C.-H., Oyang, Y.-J., Heui-Ming Ma, M., and Sun, W.-Z., 2015. Ambulance Service Area Considering Disaster-Induced Disturbance on the Transportation Infrastructure. *Journal of Testing and Evaluation*, 43 (2), 1-11.

Chen, J., Shaw, S.L., Yu, H., Lu, F., Chai, Y., and Jia, Q., 2011. Exploratory data analysis of activity diary data: A space-time GIS approach. *Journal of Transport Geography*, 19 (3), 394–404.

Chen, X. and Clark, J., 2016. Measuring space–time access to food retailers: a case of temporal access disparity in Franklin County, Ohio. *The Professional Geographer*, 68 (2), 175–188.

Chen, X. and Kwan, M.-P., 2015. Contextual Uncertainties, Human Mobility, and Perceived Food Environment: The Uncertain Geographic Context Problem in Food Access Research. *American Journal of Public Health*, 105 (9), 1734–1737.

Cheng, T. and Adepeju, M., 2014. Modifiable temporal unit problem (MTUP) and its effect on space-time cluster detection. *PLoS ONE*, 9 (6), 1–10.

Church, R.L., 1999. Location modelling and GIS. *In*: L. P, G. M, M. D, and R. D, eds. *Geographical information systems*. New York: John Wiley Chichester, Sussex, 293–303.

Clark, A. and Scott, D., 2014. Understanding the impact of the modifiable areal unit problem on the relationship between active travel and the built environment. *Urban Studies*, 51 (2), 284–299.

Cobb, L.K., Appel, L.J., Franco, M., Jones-Smith, J.C., Nur, A., and Anderson, C.A.M., 2015. The relationship of the local food environment with obesity: A systematic review of methods, study quality, and results. *Obesity*, 23 (7), 1331–1344.

Cohen, D.A., Ashwood, J.S., Scott, M.M., Overton, A., Evenson, K.R., Staten, L.K., Porter, D., McKenzie, T.L., and Catellier, D., 2006. Public parks and physical activity among adolescent girls. *Pediatrics*, 118 (5), 1381–1389.

Coker, A.L., Davis, K.E., Arias, I., Desai, S., Sanderson, M., Brandt, H.M., and Smith, P.H., 2002. Physical and mental health effects of intimate partner violence for men and women. *American Journal of Preventive Medicine*, 23 (4), 260–268.

Collins, S., 2003. Modelling spatial variations in air quality using GIS. *In: GIS And Health*. CRC Press, 99-114.

Committee on Accelerating Progress in Obesity, 2012. *Accelerating progress in obesity* *prevention: solving the weight of the nation*. Washington, D.C.: National Academies Press.

Coombes, E., Jones, A.P., and Hillsdon, M., 2010. The relationship of physical activity and overweight to objectively measured green space accessibility and use. *Social science & medicine*, 70 (6), 816–822.

Crawford, T.W., Jilcott Pitts, S.B., McGuirt, J.T., Keyserling, T.C., and Ammerman, A.S., 2014. Conceptualizing and comparing neighborhood and activity space measures for food environment research. *Health and Place*, 30, 215–225.

Cubbin, C., Santelli, J., Brindis, C.D., and Braveman, P., 2005. Neighborhood context and sexual behaviors among adolescents: findings from the national longitudinal study of adolescent health. *Perspectives on sexual and reproductive health*, 37 (3), 125–134.

Cummins, S., 2007. Commentary: investigating neighbourhood effects on health—avoiding the 'local trap'. *International Journal of Epidemiology*, 36 (2), 355–357.

Curtis, S., 2010. *Space, place and mental health*. London, UK: Routledge.

D'Amico, S.J., Wang, S.-J., Batta, R., and Rump, C.M., 2002. A simulated annealing approach to police district design. *Computers & Operations Research*, 29 (6), 667–684.

Dai, D. and Wang, F., 2011. Geographic disparities in accessibility to food stores in southwest Mississippi, 38, 659–678.

Davis, B. and Carpenter, C., 2009. Proximity of fast-food restaurants to schools and adolescent obesity. *American Journal of Public Health*, 99 (3), 505–510.

Diez Roux, A. V., Jacobs, D.R., and Kiefe, C.I., 2002. Neighborhood characteristics and components of the insulin resistance syndrome in young adults: The Coronary Artery Risk Development in Young Adults (CARDIA) Study. *Diabetes Care*, 25 (11), 1976–1982.

Diez Roux, A.V., 2001. Investigating Neighbourhood and Area Effects on Health. *American*

Journal of Public Health, 91 (11), 1783–1789.

Ding, D., Sallis, J.F., Kerr, J., Lee, S., and Rosenberg, D.E., 2011. Neighborhood environment and physical activity among youth: a review. *American journal of preventive medicine*, 41 (4), 442–455.

Dowda, M., Brown, W.H., McIver, K.L., Pfeiffer, K.A., O'Neill, J.R., Addy, C.L., and Pate, R.R., 2009. Policies and characteristics of the preschool environment and physical activity of young children. *Pediatrics*, 123 (2), e261–e266.

Duncan, M.J., Badland, H.M., and Mummery, W.K., 2009. Applying GPS to enhance understanding of transport-related physical activity. *Journal of Science and Medicine in Sport*, 12 (5), 549–556.

Duncan, S., Stewart, T.I., Oliver, M., Mavoa, S., MacRae, D., Badland, H.M., and Duncan, M.J., 2013. Portable global positioning system receivers: static validity and environmental conditions. *American journal of preventive medicine*, 44 (2), e19–e29.

Duncan, S.C., Duncan, T.E., and Strycker, L.A., 2002. A multilevel analysis of neighborhood context and youth alcohol and drug problems. *Prevention Science*, 3 (2), 125–133.

Dunn, R.A., Sharkey, J.R., and Horel, S., 2012. The effect of fast-food availability on fast-food consumption and obesity among rural residents: An analysis by race/ethnicity. *Economics and Human Biology*, 10 (1), 1–13.

Dunton, G.F., Almanza, E., Jerrett, M., Wolch, J., and Pentz, M.A., 2014. Neighborhood park use by children: Use of accelerometry and global positioning systems. *American Journal of Preventive Medicine*, 46 (2), 136–142.

Elgethun, K., Fenske, R.A., Yost, M.G., and Palcisko, G.J., 2003. Time-location analysis for exposure assessment studies of children using a novel global positioning system instrument.

Environmental health perspectives, 111 (1), 115–122.

Ellaway, A., Macintyre, S., and Bonnefoy, X., 2005. Graffiti, greenery, and obesity in adults: secondary analysis of European cross sectional survey. *Bmj*, 331 (7517), 611–612.

Entwisle, B., 2007. Putting people into place. *Demography*, 44 (4), 687–703.

Epstein, D.H., Tyburski, M., Craig, I.M., Phillips, K.A., Jobes, M.L., Vahabzadeh, M., Mezghanni, M., Lin, J.L., Furr-Holden, C.D.M., and Preston, K.L., 2014. Real-time tracking of neighborhood surroundings and mood in urban drug misusers: Application of a new method to study behavior in its geographical context. *Drug and Alcohol Dependence*, 134 (1), 22–29.

Epstein, D.H., Willner-Reid, J., Vahabzadeh, M., Mezghanni, M., Lin, J.-L., and Preston, K.L., 2009. Real-time electronic diary reports of cue exposure and mood in the hours before cocaine and heroin craving and use. *Archives of general psychiatry*, 66 (1), 88–94.

Eriksson, C., Rosenlund, M., Pershagen, G., Hilding, A., Östenson, C.-G., and Bluhm, G., 2007. Aircraft Noise and Incidence of Hypertension. *Epidemiology*, 18 (6), 716–721.

ESRI, 2017. Data classification methods [online]. Available from: http://pro.arcgis.com/en/pro-app/help/mapping/layer-properties/data-classification-methods.htm [Accessed 31 Jan 2017].

Evans, G.W., 2003. The Built Environment and Mental Health. *Journal of Urban Health*, 80 (4), 536–555.

Ewing, R. and Cervero, R., 2010. Travel and the built environment: a meta-analysis. *Journal of the American planning association*, 76 (3), 265–294.

Ewing, R., Schmid, T., Killingsworth, R., Zlot, A., and Raudenbush, S., 2003. Relationship between urban sprawl and physical activity, obesity, and morbidity. *American journal of health promotion*, 18 (1), 47–57.

Feick, R. and Robertson, C., 2015. A multi-scale approach to exploring urban places in geotagged photographs. *Computers, Environment and Urban Systems*, 53, 96–109.

Feng, J., Glass, T.A., Curriero, F.C., Stewart, W.F., and Schwartz, B.S., 2010. The built environment and obesity: a systematic review of the epidemiologic evidence. *Health & place*, 16 (2), 175–190.

Ferland, J.A. and Guénette, G., 1990. Decision Support System for the School Districting Problem. *Operations Research*, 38 (1), 15–21.

Fisher, K.J., Li, F., Michael, Y., and Cleveland, M., 2004. Neighborhood-level influences on physical activity among older adults: a multilevel analysis. *Journal of aging and physical activity*, 12 (1), 45–63.

Fleischmann, B. and Paraschis, J.N., 1988. Solving a large scale districting problem: a case report. *Computers & Operations Research*, 15 (6), 521–533.

Fontaine, K.R., Heo, M., and Bathon, J., 2004. Are US adults with arthritis meeting public health recommendations for physical activity? *Arthritis & Rheumatism*, 50 (2), 624–628.

Fowler, P.J., Tompsett, C.J., Braciszewski, J.M., Jacques-Tiura, A.J., and Baltes, B.B., 2009. Community violence: A meta-analysis on the effect of exposure and mental health outcomes of children and adolescents. *Development and Psychopathology*, 21 (1), 227–259.

Fraley, G., Jankowska, M., and Jankowski, P., 2010. Towards Memetic Algorithms in GIScience: An Adaptive Multi-Objective Algorithm for Optimized Delineation of Neighborhood Boundaries. *In*: *GIScience Conference, Zurich, Switzerland*.

Frank, L., Kerr, J., Chapman, J., and Sallis, J., 2007. Urban form relationships with walk trip frequency and distance among youth. *American journal of health promotion*, 21 (SUPPL. 4), 305–311.

Frank, L.D., Andresen, M.A., and Schmid, T.L., 2004. Obesity relationships with community design, physical activity, and time spent in cars. *American journal of preventive medicine*, 27 (2), 87–96.

Frank, L.D., Schmid, T.L., Sallis, J.F., Chapman, J., and Saelens, B.E., 2005. Linking objectively measured physical activity with objectively measured urban form: Findings from SMARTRAQ. *American Journal of Preventive Medicine*, 28 (SUPPL. 2), 117–125.

Franklin, A.D. and Koenigsberg, E., 1973. Computed School Assignments in a Large District. *Operations Research*, 21 (2), 413–426.

Franklin County Community Health Needs Assessment Steering Committee, 2013. *Franklin County HealthMap 2013*.

Gamba, R.J., Schuchter, J., Rutt, C., and Seto, E.Y.W., 2015. Measuring the Food Environment and its Effects on Obesity in the United States: A Systematic Review of Methods and Results. *Journal of Community Health*, 40 (3), 464–475.

GDAL/OGR Contributors, 2018. GDAL/OGR Geospatial Data Abstraction software Library.

George, J.A., Lamar, B.W., and Wallace, C.A., 1997. Political district determination using large-scale network optimization. *Socio-Economic Planning Sciences*, 31 (1), 11–28.

Gesler, W.M. and Meade, M.S., 1988. Locational and population factors in health care-seeking behavior in Savannah, Georgia. *Health services research*, 23 (3), 443–462.

Giles-Corti, B., Broomhall, M.H., Knuiman, M., Collins, C., Douglas, K., Ng, K., Lange, A., and Donovan, R.J., 2005. Increasing walking: how important is distance to, attractiveness, and size of public open space? *American journal of preventive medicine*, 28 (2), 169–176.

Goetz, S.J., Prince, S.D., and Small, J., 2000. Advances in satellite remote sensing of environmental variables for epidemiological applications. *Advances in Parasitology*, 47, 289–307.

Golledge, R.G., 1997. *Spatial behavior: A geographic perspective*. Guilford Press.

Gordon-Larsen, P., Nelson, M.C., and Beam, K., 2005. Associations among active transportation, physical activity, and weight status in young adults. *Obesity Research*, 13 (5), 868–875.

Gray, S.C., Edwards, S.E., and Miranda, M.L., 2013. Race, socioeconomic status, and air pollution exposure in North Carolina. *Environmental research*, 126, 152–158.

Greaves, S., Issarayangyun, T., and Liu, Q., 2008. Exploring variability in pedestrian exposure to fine particulates (PM2.5) along a busy road. *Atmospheric Environment*, 42 (8), 1665–1676.

Grow, H., Saelens, B., Kerr, J., Durant, N., Norman, G., and Sallis, J., 2008. Where are youth active? Roles of proximity, active transport, and built environment. *Medicine & Science in Sports & Exercise*, 40 (12), 2071.

Gulliver, J. and Briggs, D.J., 2005. Time-space modeling of journey-time exposure to traffic-related air pollution using GIS. *Environmental Research*, 97, 10–25.

Hakimi, S.L., 1964. Optimum Locations of Switching Centers and the Absolute Centers and Medians of a Graph. *Operations Research*, 12 (3), 450–459.

Hakimi, S.L., 1965. Optimum Distribution of Switching Centers in a Communication Network and Some Related Graph Theoretic Problems. *Operations Research*, 13 (3), 462–475.

Handy, S.L., Boarnet, M.G., Ewing, R., and Killingsworth, R.E., 2002. How the built environment affects physical activity: views from urban planning. *American journal of preventive medicine*, 23 (2), 64–73.

Hess, S.W. and Samuels, S.A., 1971. Experiences with a Sales Districting Model: Criteria and Implementation. *Management Science*, 18 (4–part–ii), 41–54.

Hess, S.W., Weaver, J.B., Siegfeldt, H.J., Whelan, J.N., and Zitlau, P.A., 1965. Nonpartisan Political Redistricting by Computer. *Operations Research*, 13 (6), 998–1006.

Hillsdon, M., Panter, J., Foster, C., and Jones, A., 2006. The relationship between access and quality of urban green space with population physical activity. *Public health*, 120 (12), 1127–1132.

Hoehner, C.M., Ramirez, L.K.B., Elliott, M.B., Handy, S.L., and Brownson, R.C., 2005. Perceived and objective environmental measures and physical activity among urban adults. *American journal of preventive medicine*, 28 (2), 105–116.

Holloway, C.A., Wehrung, D.A., Zeitlin, M.P., and Nelson, R.T., 1975. An Interactive Procedure for the School Boundary Problem with Declining Enrollment. *Operations Research*, 23 (2), 191–206.

Holsten, J.E., 2009. Obesity and the community food environment: A systematic review. *Public Health Nutrition*, 12 (3), 397–405.

Holt, J.B., 2007. The topography of poverty in the United States: a spatial analysis using county-level data from the Community Health Status Indicators project. *Preventing chronic disease*, 4 (4), A111.

Houle, J.N. and Light, M.T., 2014. The home foreclosure crisis and rising suicide rates, 2005 to 2010. *American Journal of Public Health*, 104 (6), 1073–1079.

Houston, D., 2014. Implications of the modifiable areal unit problem for assessing built environment correlates of moderate and vigorous physical activity. *Applied Geography*, 50, 40–47.

Inagami, S., Cohen, D.A., Brown, A.F., and Asch, S.M., 2009. Body mass index, neighborhood fast food and restaurant concentration, and car ownership. *Journal of Urban Health*, 86 (5), 683–695.

Inagami, S., Cohen, D.A., and Finch, B.K., 2007. Non-residential neighborhood exposures suppress neighborhood effects on self-rated health. *Social science & medicine*, 65 (8), 1779–1791.

Inagami, S., Cohen, D.A., Finch, B.K., and Asch, S.M., 2006. You are where you shop: grocery store locations, weight, and neighborhoods. *American journal of preventive medicine*, 31 (1), 10–17.

Jago, R., Baranowski, T., and Baranowski, J.C., 2006. Observed, GIS, and self-reported environmental features and adolescent physical activity. *American Journal of Health Promotion*, 20 (6), 422–428.

Jago, R., Baranowski, T., and Harris, M., 2006. Relationships between GIS environmental features and adolescent male physical activity: GIS coding differences. *Journal of Physical Activity and Health*, 3 (2), 230–242.

James, P., Berrigan, D., Hart, J.E., Aaron Hipp, J., Hoehner, C.M., Kerr, J., Major, J.M., Oka, M., and Laden, F., 2014. Effects of buffer size and shape on associations between the built environment and energy balance. *Health and Place*, 27, 162–170.

Jeffery, R.W., Baxter, J., McGuire, M., and Linde, J., 2006. Are fast food restaurants an environmental risk factor for obesity? *International Journal of Behavioral Nutrition and Physical Activity*, 3 (1), 2.

Jerrett, M., Arain, A., Kanaroglou, P., Beckerman, B., Potoglou, D., Sahsuvaroglu, T., Morrison, J., and Giovis, C., 2005. A review and evaluation of intraurban air pollution exposure models. *Journal of exposure analysis and environmental epidemiology*, 15 (2), 185–204.

Jilcott, S.B., Wade, S., McGuirt, J.T., Wu, Q., Lazorick, S., and Moore, J.B., 2011. The association between the food environment and weight status among eastern North Carolina youth. *Public Health Nutrition*, 14 (9), 1610–1617.

Jones, A.P., Coombes, E.G., Griffin, S.J., and van Sluijs, E.M., 2009. Environmental supportiveness for physical activity in English schoolchildren: a study using Global Positioning Systems. *International Journal of Behavioral Nutrition and Physical Activity*, 6 (1), 42.

Kerr, J., Frank, L., Sallis, J.F., and Chapman, J., 2007. Urban form correlates of pedestrian travel in youth: Differences by gender, race-ethnicity and household attributes. *Transportation Research Part D: Transport and Environment*, 12 (3), 177–182.

Kerr, J., Norman, G.J., Sallis, J.F., and Patrick, K., 2008. Exercise aids, neighborhood safety, and physical activity in adolescents and parents. *Medicine and science in sports and exercise*, 40 (7), 1244–1248.

Kestens, Y., Lebel, A., Chaix, B., Clary, C., Daniel, M., Pampalon, R., Theriault, M., and Subramanian, S. V., 2012. Association between activity space exposure to food establishments and individual risk of overweight. *PLoS ONE*, 7 (8), e41418.

Kestens, Y., Thierry, B., and Chaix, B., 2016. Re-creating daily mobility histories for health research from raw GPS tracks: Validation of a kernel-based algorithm using real-life data. *Health and Place*, 40, 29–33.

Kestens, Y., Thierry, B., Shareck, M., Steinmetz-Wood, M., and Chaix, B., 2018. Integrating activity spaces in health research: Comparing the VERITAS activity space questionnaire with 7-day GPS tracking and prompted recall. *Spatial and Spatio-temporal Epidemiology*, 25, 1–9.

Kestens, Y., Wasfi, R., Naud, A., and Chaix, B., 2017. “Contextualizing Context”: Reconciling

Environmental Exposures, Social Networks, and Location Preferences in Health Research. *Current Environmental Health Reports*, 4 (1), 51–60.

King, A.C., Castro, C., Wilcox, S., Eyler, A.A., Sallis, J.F., and Brownson, R.C., 2000. Personal and environmental factors associated with physical inactivity among different racial-ethnic groups of US middle-aged and older-aged women. *Health psychology*, 19 (4), 354.

Koohsari, M.J., Badland, H., and Giles-Corti, B., 2013. (Re)Designing the built environment to support physical activity: Bringing public health back into urban design and planning. *Cities*, 35, 294–298.

Koohsari, M.J., Mavoa, S., Villianueva, K., Sugiyama, T., Badland, H., Kaczynski, A.T., Owen, N., and Giles-Corti, B., 2015. Public open space, physical activity, urban design and public health: Concepts, methods and research agenda. *Health and Place*, 33, 75–82.

Krause, C., 2013. An Activity Space Based Approach for Capturing Long Distance Travel Using Longitudinal GPS Survey Data. *In: Transportation Research Board 92nd annual meeting*. Washington DC. Paper (No. 13-3079).

Krenn, P.J., Titze, S., Oja, P., Jones, A., and Ogilvie, D., 2011. Use of global positioning systems to study physical activity and the environment: A systematic review. *American Journal of Preventive Medicine*, 41 (5), 508–515.

Kurka, J.M., Adams, M.A., Todd, M., Colburn, T., Sallis, J.F., Cain, K.L., Glanz, K., Frank, L.D., and Saelens, B.E., 2015. Patterns of neighborhood environment attributes in relation to children’s physical activity. *Health & place*, 34, 164–170.

Kwan, M.-P., 2000. Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: a methodological exploration with a large data set. *Transportation Research Part C: Emerging Technologies*, 8 (1), 185–203.

Kwan, M.-P., 2009a. From place-based to people-based exposure measures. *Social Science & Medicine*, 69 (9), 1311–1313.

Kwan, M.-P., 2013. Beyond space (as we knew it): toward temporally integrated geographies of segregation, health, and accessibility: Space–time integration in geography and GIScience. *Annals of the Association of American Geographers*, 103 (5), 1078–1086.

Kwan, M.-P., 2018a. The Neighborhood Effect Averaging Problem (NEAP): An Elusive Confounder of the Neighborhood Effect. *International Journal of Environmental Research and Public Health*, 15 (9), 1841.

Kwan, M.-P., 2012a. The Uncertain Geographic Context Problem. *Annals of the Association of American* Geographers, 102 (5), 958–968.

Kwan, M.-P., 2012b. How GIS can help address the uncertain geographic context problem in social science research. *Annals of GIS*, 18 (4), 245–255.

Kwan, M.-P., Wang, J., Tyburski, M., Epstein, D.H., Kowalczyk, W.J., and Preston, K.L.,. Uncertainties in the geographic context of health behaviors: a study of substance users ' exposure to psychosocial stress using GPS data. *International Journal of Geographical Information Science* (forthcoming). doi:10.1080/13658816.2018.1503276.

Kwan, M.-P. and Weber, J., 2003. Individual accessibility revisited: implications for geographical analysis in the twenty-first century. *Geographical analysis*, 35 (4), 341–353.

Kwan, M.-P., 2009b. From place-based to people-based exposure measures. *Social Science and Medicine*, 69 (9), 1311–1313.

Kwan, M.-P., 2018b. The Limits of the Neighborhood Effect: Contextual Uncertainties in Geographic, Environmental Health, and Social Science Research. *Annals of the American Association of Geographers*, 108(6), 1482-1490.

Kwan, M.-P., Kenda, L.L., Wewers, M.E., Ferketich, A.K., and Klein, E.G., 2011. Sociogeographic context, protobacco advertising, and smokeless tobacco usage in the Appalachian Region of Ohio (USA). *In*: *International Medical Geography Symposium*. Durham, UK.

Laatikainen, T.E., Hasanzadeh, K., and Kyttä, M., 2018. Capturing exposure in environmental health research: challenges and opportunities of different activity space models. *International Journal of Health Geographics*, 17 (1), 29.

Lachowycz, K. and Jones, A.P., 2011. Greenspace and obesity: A systematic review of the evidence. *Obesity Reviews*, 12 (501), 183–189.

Lachowycz, K., Jones, A.P., Page, A.S., Wheeler, B.W., and Cooper, A.R., 2012. What can global positioning systems tell us about the contribution of different types of urban greenspace to children's physical activity? *Health and Place*, 18 (3), 586–594.

Lahti, J., Holstila, A., Lahelma, E., and Rahkonen, O., 2014. Leisure-time physical activity and all-cause mortality. *PLoS ONE*, 9 (7), e101548.

Lamichhane, A.P., Puett, R., Porter, D.E., Bottai, M., Mayer-Davis, E.J., and Liese, A.D., 2012. Associations of built food environment with body mass index and waist circumference among youth with diabetes. *International Journal of Behavioral Nutrition and Physical Activity*, 9 (1), 81.

Larsen, K., Gilliland, J., Hess, P., Tucker, P., Irwin, J., and He, M., 2009. The influence of the physical environment and sociodemographic characteristics on children's mode of travel to and from school. *American Journal of Public Health*, 99 (3), 520–526.

Leal, C. and Chaix, B., 2011. The influence of geographic life environments on cardiometabolic risk factors: a systematic review, a methodological assessment and a research agenda.

Obesity Reviews, 12 (3), 217–230.

Lee, B.A., Reardon, S.F., Firebaugh, G., Farrell, C.R., Matthews, S.A., and O'Sullivan, D., 2008. Beyond the Census Tract: Patterns and Determinants of Racial Segregation at Multiple Geographic Scales. *American Sociological Review*, 73 (5), 766–791.

Lee, C. and Moudon, A.V., 2006. Correlates of walking for transportation or recreation purposes. *Journal of Physical Activity & Health*, 3 (s1), S77–S98.

Lee, H., 2012. The role of local food availability in explaining obesity risk among young school-aged children. *Social Science and Medicine*, 74 (8), 1193–1203.

Lee, M., Brown, A., and Goodchild, M., 2013. Does distance decay modelling of supermarket accessibility predict fruit and vegetable intake by individuals in a large metropolitan area? *Journal of Health Care for the Poor and Underserved*, 24 (10), 172–185.

Lee, R.E. and Cubbin, C., 2002. Neighborhood context and youth cardiovascular health behaviors. *American Journal of Public Health*, 92 (3), 428–436.

Li, F., Harmer, P., Cardinal, B.J., Bosworth, M., and Johnson-Shelton, D., 2009. Obesity and the built environment: Does the density of neighborhood fast-food outlets matter? *American Journal of Health Promotion*, 23 (3), 203–209.

Liao, Y., Intille, S.S., and Dunton, G.F., 2015. Using ecological momentary assessment to understand where and with whom adults' physical and sedentary activity occur. *International journal of behavioral medicine*, 22 (1), 51–61.

Lin, C. and Huang, I., 1998. A Crystal Growth Approach for Topographical Global Optimization. *Journal of Global Optimization*, 13 (3), 255–267.

Lipperman-Kreda, S., Morrison, C., Grube, J.W., and Gaidus, A., 2015. Youth activity spaces and daily exposure to tobacco outlets. *Health and Place*, 34, 30–33.

Lobitz, B., Beck, L., Huq, A., Wood, B., Fuchs, G., Faruque, A.S.G., and Colwell, R., 2000. Climate and infectious disease: use of remote sensing for detection of Vibrio cholerae by indirect measurement. *Proceedings of the National Academy of Sciences*, 97 (4), 1438–1443.

Lovasi, G.S., Grady, S., and Rundle, A., 2012. Steps Forward: Review and Recommendations for Research on Walkability, Physical Activity and Cardiovascular Health. *Public Health Rev*, 33 (4), 484–506.

Love, R.F., 1972. A computational procedure for optimally locating a facility with respect to several rectangular regions. *Journal of Regional Science*, 12 (2), 233–242.

Lukasiewicz, M., Fareng, M., Benyamina, A., Blecha, L., Reynaud, M., and Falissard, B., 2007. Ecological momentary assessment in addiction. *Expert review of neurotherapeutics*, 7 (8), 939–950.

Lytle, L.A. and Sokol, R.L., 2017. Measures of the food environment: A systematic review of the field, 2007–2015. *Health and Place*, 44, 18–34.

Ma, H., Carlin, B.P., and Banerjee, S., 2010. Hierarchical and Joint Site-Edge Methods for Medicare Hospice Service Region Boundary Analysis. *Biometrics*, 66 (2), 355–364.

Maantay, J.A. and McLafferty, S., 2011. Environmental health and geospatial analysis: an overview. *In*: *Geospatial Analysis of Environmental Health*. Springer, Dordrecht, 3–37.

Maas, J., Verheij, R.A., Spreeuwenberg, P., and Groenewegen, P.P., 2008. Physical activity as a possible mechanism behind the relationship between green space and health: a multilevel analysis. *BMC public health*, 8 (1), 206.

Macintyre, S., Ellaway, A., and Cummins, S., 2002. Place effects on health: how can we conseptualise, operationalise and measure them? *Social Science and Medicine*, 55, 125–139.

Macmillan, W., 2001. Redistricting in a GIS environment: An optimisation algorithm using switching-points. *Journal of Geographical Systems*, 3 (2), 167–180.

Maddison, R. and Ni Mhurchu, C., 2009. Global positioning system: a new opportunity in physical activity measurement. *The international journal of behavioral nutrition and physical activity*, 6, 73.

Maddock, J., 2004. The relationship between obesity and the prevalence of fast food restaurants: state-level analysis. *American journal of health promotion*, 19 (2), 137–143.

Marlin, P.G., 1981. Application of the transportation model to a large-scale "Districting" problem. *Computers & Operations Research*, 8 (2), 83–96.

Maroko, A.R., Maantay, J.A., Sohler, N.L., Grady, K.L., and Arno, P.S., 2009. The complexities of measuring access to parks and physical activity sites in New York City: a quantitative and qualitative approach. *International journal of health geographics*, 8 (1), 34.

Matthews, S.A., 2008. The salience of neighborhood: some lessons from sociology. *American journal of preventive medicine*, 34 (3), 257–259.

McGinn, A.P., Evenson, K.R., Herring, A.H., Huston, S.L., and Rodriguez, D.A., 2007. Exploring associations between physical activity and perceived and objective measures of the built environment. *Journal of Urban Health*, 84 (2), 162–184.

Mellor, J.M., Dolan, C.B., and Rapoport, R.B., 2011. Child body mass index, obesity, and proximity to fast food restaurants. *International Journal of Pediatric Obesity*, 6 (1), 60–68.

Melnick, A.L., 2002. *Introduction to geographic information systems in public health*. Jones & Bartlett Learning.

Mennis, J. and Mason, M.J., 2011. People, Places, and Adolescent Substance Use: Integrating Activity Space and Social Network Data for Analyzing Health Behavior. *Annals of the*

Association of American Geographers, 101 (2), 272–291.

Miller, H., 2007. Place-Based versus People-Based Geographic Information Science. *Geography Compass*, 1 (3), 503–535.

Miller, H.J., 1996. GIS and geometric representation in facility location problems. *International Journal of Geographical Information Systems*, 10 (7), 791–816.

Millstein, R.A., Yeh, H.-C., Brancati, F.L., Batts-Turner, M., and Gary, T.L., 2009. Food availability, neighborhood socioeconomic status, and dietary patterns among blacks with type 2 diabetes mellitus. *The Medscape Journal of Medicine*, 11 (1), 15.

Mitchell, C.A., Clark, A.F., and Gilliland, J.A., 2016. Built environment influences of children's physical activity: Examining differences by neighbourhood size and sex. *International Journal of Environmental Research and Public Health*, 13 (1), 130.

Mitra, R. and Buliung, R.N., 2012. Built environment correlates of active school transportation: neighborhood and the modifiable areal unit problem. *Journal of transport geography*, 20 (1), 51–61.

Mokdad, A.H., Marks, J.S., Stroup, D.F., and Gerberding, J.L., 2004. Actual causes of death in the United States, 2000. *Jama*, 291 (10), 1238–1245.

Moore, K., Roux, A.V.D., Auchincloss, A., Evenson, K.R., Kaufman, J., Mujahid, M., and Williams, K., 2013. Home and work neighbourhood environments in relation to body mass index : the Multi-Ethnic Study of Atherosclerosis (MESA). *J Epidemiol Community Health*, 67 (10), 846–853.

Moore, S.C., Lee, I.M., Weiderpass, E., Campbell, P.T., Sampson, J.N., Kitahara, C.M., Keadle, S.K., Arem, H., De Gonzalez, A.B., Hartge, P., Adami, H.O., Blair, C.K., Borch, K.B., Boyd, E., Check, D.P., Fournier, A., Freedman, N.D., Gunter, M., Johannson, M., Khaw, K.T., Linet, M.S., Orsini, N., Park, Y., Riboli, E., Robien, K., Schairer, C., Sesso, H., Spriggs, M., Van Dusen, R., Wolk, A., Matthews, C.E., and Patel, A. V., 2016. Association of leisure-time physical activity with risk of 26 types of cancer in 1.44 million adults. *JAMA Internal Medicine*, 176 (6), 816–825.

Moreno-Regidor, P., García López de Lacalle, J., and Manso-Callejo, M.-Á., 2012. Zone design of specific sizes using adaptive additively weighted Voronoi diagrams. *International Journal of Geographical Information Science*, 26 (10), 1811–1829.

Morenoff, J.D., House, J.S., Hansen, B.B., Williams, D.R., Kaplan, G.A., and Hunte, H.E., 2007. Understanding social disparities in hypertension prevalence, awareness, treatment, and control: The role of neighborhood context. *Social Science and Medicine*, 65 (9), 1853–1866.

Morland, K.B. and Evenson, K.R., 2009. Obesity prevalence and the local food environment. *Health and Place*, 15 (2), 491–495.

Mota, J., Almeida, M., Santos, P., and Ribeiro, J.C., 2005. Perceived neighborhood environments and physical activity in adolescents. *Preventive medicine*, 41 (5), 834–836.

Mu, L. and Wang, F., 2008. A scale-space clustering method: Mitigating the effect of scale in the analysis of zone-based data. *Annals of the Association of American Geographers*, 98 (1), 85–101.

Murray, A.T., 2003. Site placement uncertainty in location analysis. *Computers, Environment and Urban Systems*, 27 (2), 205–221.

Murray, A.T., 2008. Geography in coverage modelling: exploiting spatial strucutre to address complementray partial service of areas. *Annals of the Association of American Geographers*, 95 (4), 761–772.

Murray, A.T., O'Kelly, M.E., and Church, R.L., 2008. Regional service coverage modeling.

Computers and Operations Research, 35 (2), 339–355.

Nagel, C.L., Carlson, N.E., Bosworth, M., and Michael, Y.L., 2008. The relation between neighborhood built environment and walking activity among older adults. *American journal of epidemiology*, 168 (4), 461–468.

Nelson, M.C., Gordon-Larsen, P., Song, Y., and Popkin, B.M., 2006. Built and Social Environments. Associations with Adolescent Overweight and Activity. *American Journal of Preventive Medicine*, 31 (2), 109–117.

Nyhan, M., Grauwin, S., Britter, R., Misstear, B., McNabola, A., Laden, F., Barrett, S.R.H., and Ratti, C., 2016. "Exposure Track" The Impact of Mobile-Device-Based Mobility Patterns on Quantifying Population Exposure to Air Pollution. *Environmental science & technology*, 50 (17), 9671–9681.

O'Donovan, G., Lee, I.M., Hamer, M., and Stamatakis, E., 2017. Association of 'weekend warrior' and other leisure time physical activity patterns with risks for all-cause, cardiovascular disease, and cancer mortality. *JAMA Internal Medicine*, 177 (3), 335–342.

Oakes, J.M., Forsyth, A., and Schmitz, K.H., 2007. The effects of neighborhood density and street connectivity on walking behavior: the Twin Cities walking study. *Epidemiologic perspectives & innovations*, 4 (1), 16.

Ogden, C.L., Carroll, M.D., Kit, B.K., and Flegal, K.M., 2014. Prevalence of childhood and adult obesity in the United States, 2011-2012. *Jama*, 311 (8), 806–814.

Okabe, A. and Suzuki, A., 1997. Locational optimization problems solved through Voronoi diagrams. *European Journal of Operational Research*, 98 (3), 445–456.

Oliver, L.N. and Hayes, M. V, 2008. Effects of neighbourhood income on reported body mass index: an eight year longitudinal study of Canadian children. *BMC Public Health*, 8 (1), 16.

Oliver, L.N., Schuurman, N., and Hall, A.W., 2007. Comparing circular and network buffers to examine the influence of land use on walking for leisure and errands. *International journal of health geographics*, 6 (1), 41.

Openshaw, S., 1984. The modifiable areal unit problem. *Concepts and techniques in modern geography*.

Openshaw, S., 1996. Developing GIS-relevant zone-based spatial analysis methods. *In: Spatial analysis: modelling in a GIS environment.* John Wiley & Sons , 55–73.

Openshaw, S. and Rao, L., 1995. Algorithms for Reengineering 1991 Census Geography. *Environment and Planning A*, 27 (3), 425–446.

Páez, A., Mercado, R.G., Farber, S., Morency, C., and Roorda, M., 2010. Relative Accessibility Deprivation Indicators for Urban Settings : Definitions and Application to Food Deserts in Montreal. *Urban Studies*, 47 (7), 1415–1438.

PAGAC, 2008. *Physical Activity Guidelines Advisory Committee report*. Washington, DC.

Park, Y.M. and Kwan, M.-P., 2017. Individual exposure estimates may be erroneous when spatiotemporal variability of air pollution and human mobility are ignored. *Health and Place*, 43, 85–94.

Pearce, J., 2000. Techniques for defining school catchment areas for comparison with census data. *Computers, Environment and Urban Systems*, 24 (4), 283–303.

Perchoux, C., Chaix, B., Cummins, S., and Kestens, Y., 2013. Conceptualization and measurement of environmental exposure in epidemiology: Accounting for activity space related to daily mobility. *Health and Place*, 21, 86–93.

Phillips, M.L., Hall, T.A., Esmen, N.A., Lynch, R., and Johnson, D.L., 2001. Use of global positioning system technology to track subject's location during environmental exposure sampling. *Journal of exposure analysis and environmental epidemiology*, 11 (3), 207–215.

Prüss-Ustün, A., Wolf, J., Corvalán, C., Bos, R., and Neira, M., 2016. *Preventing disease through healthy environments: a global assessment of the burden of disease from environmental risks*. Geneva, Switzerland.

Quigg, R., Gray, A., Reeder, A.I., Holt, A., and Waters, D.L., 2010. Using accelerometers and GPS units to identify the proportion of daily physical activity located in parks with playgrounds in New Zealand children. *Preventive medicine*, 50 (5–6), 235–240.

Rainham, D., McDowell, I., Krewski, D., and Sawada, M., 2010. Conceptualizing the healthscape: Contributions of time geography, location technologies and spatial ecology to place and health research. *Social Science and Medicine*, 70 (5), 668–676.

Rainham, D.G., Bates, C.J., Blanchard, C.M., Dummer, T.J., Kirk, S.F., and Shearer, C.L., 2012. Spatial classification of youth physical activity patterns. *American journal of preventive medicine*, 42 (5), e87–e96.

Reisig, V. and Hobbiss, A., 2000. Food deserts and how to tackle them: a study of one city's approach. *Health Education Journal*, 59, 137–149.

Ricca, F., Scozzari, A., and Simeone, B., 2008. Weighted Voronoi region algorithms for political districting. *Mathematical and Computer Modelling*, 48 (9–10), 1468–1477.

Ricca, F., Scozzari, A., and Simeone, B., 2013. Political districting: from classical models to recent approaches. *Annals of Operations Research*, 204 (1), 271–299.

Ríos-Mercado, R.Z. and Fernández, E., 2009. A reactive GRASP for a commercial territory design problem with multiple balancing requirements. *Computers & Operations Research*, 36 (3), 755–776.

Roberts, H.V., 2017. Using Twitter data in urban green space research: A case study and critical evaluation. *Applied Geography*, 81, 13–20.

Rodríguez, D. a., Cho, G.-H., Evenson, K.R., Conway, T.L., Cohen, D., Ghosh-Dastidar, B., Pickrel, J.L., Veblen-Mortenson, S., and Lytle, L. a., 2012. Out and about: Association of the built environment with physical activity behaviors of adolescent females. *Health & Place*, 18 (1), 55–62.

Rodriguez, D.A., Brown, A.L., and Troped, P.J., 2005. Portable global positioning units to complement accelerometry-based physical activity monitors. *Medicine and Science in Sports and Exercise*, 37 (SUPPL. 11), S572–S581.

Roemmich, J.N., Epstein, L.H., Raja, S., and Yin, L., 2007. The neighborhood and home environments: disparate relationships with physical activity and sedentary behaviors in youth. *Annals of Behavioral Medicine*, 33 (1), 29–38.

Roemmich, J.N., Epstein, L.H., Raja, S., Yin, L., Robinson, J., and Winiewicz, D., 2006. Association of access to parks and recreational facilities with the physical activity of young children. *Preventive medicine*, 43 (6), 437–441.

Rogers, D.J., Hay, S.I., and Packer, M.J., 1996. Predicting the distribution of tsetse flies in West Africa using temporal Fourier processed meteorological satellite data. *Annals of Tropical Medicine & Parasitology*, 90 (3), 225–241.

Ross, C.E., 2000a. Walking, exercising, and smoking: does neighborhood matter? *Social science & medicine*, 51 (2), 265–274.

Ross, C.E., 2000b. Neighborhood Disadvantage and Adult Depression. *Journal of Health and Social Behavior*, 41 (2), 177–187.

Rundle, A., Neckerman, K.M., Freeman, L., Lovasi, G.S., Purciel, M., Quinn, J., Richards, C., Sircar, N., and Weiss, C., 2009. Neighborhood food environment and walkability predict obesity in New York City. *Environmental Health Perspectives*, 117 (3), 442–447.

Saarloos, D., Kim, J.-E., and Timmermans, H., 2009. The built environment and health: Introducing individual space-time behavior. *International Journal of Environmental Research and Public Health*, 6 (6), 1724–1743.

Saelens, B.E., Sallis, J.F., Black, J.B., and Chen, D., 2003. Neighborhood-based differences in physical activity: an environment scale evaluation. *American journal of public health*, 93 (9), 1552–1558.

Saelens, B.E., Sallis, J.F., and Frank, L.D., 2003. Environmental correlates of walking and cycling: findings from the transportation, urban design, and planning literatures. *Annals of behavioral medicine*, 25 (2), 80–91.

Sahr, K., White, D., and Kimerling, a. J., 2003. Geodesic Discrete Global Grid Systems. *Cartography and Geographic Information Science*, 30 (2), 121–134.

Sallis, J., Bauman, A., and Pratt, M., 1998. Environmental and policy interventions to promote physical activity. *American journal of preventive medicine*, 15 (4), 379–397.

Sallis, J.F., Bowles, H.R., Bauman, A., Ainsworth, B.E., Bull, F.C., Craig, C.L., Sjöström, M., De Bourdeaudhuij, I., Lefevre, J., and Matsudo, V., 2009. Neighborhood environments and physical activity among adults in 11 countries. *American journal of preventive medicine*, 36 (6), 484–490.

Sallis, J.F., Cerin, E., Conway, T.L., Adams, M.A., Frank, L.D., Pratt, M., Salvo, D., Schipperijn, J., Smith, G., Cain, K.L., and Davey, R., 2016. Physical activity in relation to urban environments in 14 cities worldwide: a cross-sectional study. *The Lancet*, 387 (10034), 2207–2217.

Sallis, J.F., Cervero, R.B., Ascher, W., Henderson, K.A., Kraft, M.K., and Kerr, J., 2006. An ecological approach to creating active living communities. *Annual Review of Public Health*, 27, 297–322.

Salmon, J., Owen, N., Bauman, A., Schmitz, M.K.H., and Booth, M., 2000. Leisure-time, occupational, and household physical activity among professional, skilled, and less-skilled workers and homemakers. *Preventive medicine*, 30 (3), 191–199.

Santos, M.P., Page, A.S., Cooper, A.R., Ribeiro, J.C., and Mota, J., 2009. Perceptions of the built environment in relation to physical activity in Portuguese adolescents. *Health & Place*, 15 (2), 548–552.

Schaudt, B.F. and Drysdale, R.L.S., 1991. Multiplicatively weighted crystal growth Voronoi diagrams (extended abstract). *In*: *Proceedings of the seventh annual symposium on Computational geometry - SCG '91*. New York, New York, USA: ACM Press, 214–223.

Schlecht, E., Hülsebusch, C., Mahler, F., and Becker, K., 2004. The use of differentially corrected global positioning system to monitor activities of cattle at pasture. *Applied Animal Behaviour Science*, 85 (3), 185–202.

Schönfelder, S. and Axhausen, K.W., 2003. Activity spaces: Measures of social exclusion? *Transport Policy*, 10 (4), 273–286.

Schootman, M., Andresen, E.M., Wolinsky, F.D., Malmstrom, T.K., Miller, J.P., Yan, Y., and Miller, D.K., 2007. The effect of adverse housing and neighborhood conditions on the development of diabetes mellitus among middle-aged African Americans. *American Journal of Epidemiology*, 166 (4), 379–387.

Schwanen, T. and Wang, D., 2014. Well-Being, Context, and Everyday Activities in Space and Time. *Annals of the Association of American Geographers*, 104 (4), 833–851.

Seliske, L.M., Pickett, W., Boyce, W.F., and Janssen, I., 2009. Association between the food retail environment surrounding schools and overweight in Canadian youth. *Public Health Nutrition*, 12 (09), 1384.

Serre, F., Fatseas, M., Swendsen, J., and Auriacombe, M., 2015. Ecological momentary assessment in the investigation of craving and substance use in daily life: a systematic review. *Drug & Alcohol Dependence*, 148, 1–20.

Shanker, R.J., Turner, R.E., and Zoltners, A.A., 1975. Sales Territory Design: An Integrated Approach. *Management Science*, 22 (3), 309–320.

Shannon, G.W. and Spurlock, C.W., 1976. Urban Ecological Containers, Environmental Risk Cells, and the Use of Medical Services. *Economic Geography*, 52 (2), 171–180.

Shareck, M., Kestens, Y., and Frohlich, K.L., 2014. Moving beyond the residential neighborhood to explore social inequalities in exposure to area-level disadvantage: Results from the Interdisciplinary Study on Inequalities in Smoking. *Social Science & Medicine*, 108, 106–114.

Shareck, M., Kestens, Y., Vallée, J., Datta, G., Frohlich, K.L., Vallee, J., Datta, G., and Frohlich, K.L., 2015. The added value of accounting for activity space when examining the association between tobacco retailer availability and smoking among young adults. *Tobacco Control*, 25 (1), 1–7.

Sharp, G., Denney, J.T., and Kimbro, R.T., 2015. Multiple contexts of exposure: Activity spaces, residential neighborhoods, and self-rated health. *Social Science and Medicine*, 146, 204–213.

Shen, Y., Chai, Y., and Kwan, M.P., 2015. Space-time fixity and flexibility of daily activities and the built environment: A case study of different types of communities in Beijing suburbs. *Journal of Transport Geography*, 47, 90–99.

Shen, Y. and Chai, Y.W., 2013. Daily activity space of suburban mega-community residents in Beijing based on GPS data. *Acta Geographica Sinica*, 68 (4), 506–516.

Shen, Y., Kwan, M.P., and Chai, Y., 2013. Investigating commuting flexibility with GPS data and 3D geovisualization: A case study of Beijing, China. *Journal of Transport Geography*, 32, 1–11.

Sherman, J.E., Spencer, J., Preisser, J.S., Gesler, W.M., and Arcury, T.A., 2005. A suite of methods for representing activity space in a healthcare accessibility study. *International Journal of Health Geographics*, 4, 1–21.

Shiffman, S., 2009. Ecological Momentary Assessment (EMA) in Studies of Substance Use. *Psychological Assessment*, 21 (4), 486–497.

Shiffman, S., Balabanis, M.H., Gwaltney, C.J., Paty, J.A., Gnys, M., Kassel, J.D., Hickcox, M., and Paton, S.M., 2007. Prediction of lapse from associations between smoking and situational antecedents assessed by ecological momentary assessment. *Drug and Alcohol Dependence*, 91 (2–3), 159–168.

Shiffman, S., Gwaltney, C.J., Balabanis, M.H., Liu, K.S., Paty, J.A., Kassel, J.D., Hickcox, M., and Gnys, M., 2002. Immediate antecedents of cigarette smoking: An analysis from ecological momentary assessment. *Journal of Abnormal Psychology*, 111 (4), 531–545.

Shiffman, S., Stone, A.A., and Hufford, M.R., 2008. Ecological Momentary Assessment. *Annual Review of Clinical Psychology*, 4 (1), 1–32.

Shirabe, T., 2005. A Model of Contiguity for Spatial Unit Allocation. *Geographical Analysis*, 37 (1), 2–16.

Sinha, P. and Zoltners, A.A., 2001. Sales-Force Decision Models: Insights from 25 Years of Implementation. *Interfaces*, 31 (SUPPL. 3), S8–S44.

Siordia, C., Saenz, J., and Tom, S.E., 2012. An introduction to macro-level spatial nonstationarity: a geographically weighted regression analysis of diabetes and poverty. *Human geographies*, 6 (2), 5.

Spielman, S. and Yoo, E., 2009. The spatial dimensions of neighborhood effects. *Social Science & Medicine*, 68 (6), 1098–105.

Steiner, M.T.A., Datta, D., Steiner Neto, P.J., Scarpin, C.T., and Rui Figueira, J., 2015. Multi-objective optimization in partitioning the healthcare system of parana state in brazil. *Omega (United Kingdom)*, 52, 53–64.

Stigsdotter, U.K., Ekholm, O., Schipperijn, J., Toftager, M., Kamper-Jorgensen, F., and Randrup, T.B., 2010. Health promoting outdoor environments-Associations between green space, and health, health-related quality of life and stress based on a Danish national representative survey. *Scandinavian Journal of Social Medicine*, 38 (4), 411–417.

Stone, A.A. and Shiffman, S., 1994. Ecological momentary assessment (EMA) in behavorial medicine. *Annals of Behavioral Medicine*, 16(3), 199-202.

Suzuki, A. and Drezner, Z., 1996. The p-center location problem in an area. *Location science*, 4 (1–2), 69–82.

Ta, N., Chai, Y., and Kwan, M.-P., 2015. Suburbanization, daily lifestyle and space-behavior interaction: A study of suburban residents in Beijing, China. *Acta Geographica Sinica*, 70 (8), 1271–1280.

Tamura, K., Elbel, B., Chaix, B., Regan, S.D., Al-Ajlouni, Y.A., Athens, J.K., Meline, J., and Duncan, D.T., 2017. Residential and GPS-Defined Activity Space Neighborhood Noise Complaints, Body Mass Index and Blood Pressure Among Low-Income Housing Residents in New York City. *Journal of Community Health*, 42 (5), 974–982.

Tana, Kwan, M.P., and Chai, Y., 2016. Urban form, car ownership and activity space in inner suburbs: A comparison between Beijing (China) and Chicago (United States). *Urban Studies*, 53 (9), 1784–1802.

Tatem, A.J. and Hay, S.I., 2004. Measuring urbanization pattern and extent for malaria research: a review of remote sensing approaches. *Journal of Urban Health*, 81 (3), 363–376.

Thacker, S.B., Stroup, D.F., Parrish, G., and Anderson, H.A., 1996. Surveillance in environmental public health: Issues, systems, and sources. *American Journal of Public Health*, 86 (5), 633–638.

Tilt, J.H., Unfried, T.M., and Roca, B., 2007. Using objective and subjective measures of neighborhood greenness and accessible destinations for understanding walking trips and BMI in Seattle, Washington. *American Journal of Health Promotion*, 21 (SUPPL. 4), 371–379.

Timperio, A., Giles-Corti, B., Crawford, D., Andrianopoulos, N., Ball, K., Salmon, J., and Hume, C., 2008. Features of public open spaces and physical activity among children: findings from the CLAN study. *Preventive medicine*, 47 (5), 514–518.

Tobler, W.R., 1989. Frame independent spatial analysis. *In: The Accuracy of spatial databases*. CRC Press, 115–122.

Troped, P.J., Saunders, R.P., Pate, R.R., Reininger, B., and Addy, C.L., 2003. Correlates of recreational and transportation physical activity among adults in a New England community. *Preventive medicine*, 37 (4), 304–310.

Troped, P.J., Wilson, J.S., Matthews, C.E., Cromley, E.K., and Melly, S.J., 2010. The Built Environment and Location-Based Physical Activity. *American Journal of Preventive Medicine*, 38 (4), 429–438.

Trost, S.G., Owen, N., Bauman, A.E., Sallis, J.F., and Brown, W., 2002. Correlates of adults' participation in physical activity: review and update. *Medicine and science in sports and exercise*, 34 (12), 1996–2001.

Tucker, P., Irwin, J.D., Gilliland, J., He, M., Larsen, K., and Hess, P., 2009. Environmental influences on physical activity levels in youth. *Health & Place*, 15 (1), 357–363.

Vallée, J., Cadot, E., Grillo, F., Parizot, I., and Chauvin, P., 2010. The combined effects of activity space and neighbourhood of residence on participation in preventive health-care activities: The case of cervical screening in the Paris metropolitan area (France). *Health and Place*, 16 (5), 838–852.

Vandevijvere, S., Dominick, C., and Swinburn, B., 2015. The healthy food environment policy index: findings of an expert panel in New Zealand. *Bulletin of the World Health Organization*, 93 (5), 294–302.

Voronoï, G., 1908. Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Deuxième mémoire. Recherches sur les parallélloèdres primitifs. *Journal für die reine und angewandte Mathematik*, 134, 198–287.

Walker, R.E., Keane, C.R., and Burke, J.G., 2010. Disparities and access to healthy food in the United States: A review of food deserts literature. *Health and Place*, 16 (5), 876–884.

Wang, J., Kwan, M.-P., and Chai, Y., 2018. An Innovative Context-Based Crystal-Growth Activity Space Method for Environmental Exposure Assessment: A Study Using GIS and GPS Trajectory Data Collected in Chicago. *International Journal of Environmental Research and Public Health*, 15 (4), 703–726.

Wang, J., Kwan, M.-P., and Ma, L., 2014. Delimiting service area using adaptive crystal-growth Voronoi diagrams based on weighted planes: A case study in Haizhu District of Guangzhou in China. *Applied Geography*, 50 (June), 108–119.

Wang, J., Lee, K., and Kwan, M.-P., 2018. Environmental Influences on Leisure-Time Physical Inactivity in the U.S.: An Exploration of Spatial Non-Stationarity. *ISPRS International Journal of Geo-Information*, 7 (4), 143–165.

Weber, J. and Kwan, M.-P., 2003. Evaluating the Effects of Geographic Contexts on Individual Accessibility: A Multilevel Approach. *Urban Geography*, 24 (8), 647–671.

Wheaton, B. and Clarke, P., 2016. Space Meets Time: Integrating Temporal and Contextual Influences on Mental Health in Early Adulthood, 68 (5), 680–706.

Wheeler, B.W., Cooper, A.R., Page, A.S., and Jago, R., 2010. Greenspace and children's physical activity: A GPS/GIS analysis of the PEACH project. *Preventive Medicine*, 51 (2), 148–152.

Wiehe, S.E., Carroll, A.E., Liu, G.C., Haberkorn, K.L., Hoch, S.C., Wilson, J.S., and Fortenberry, J.D., 2008. Using GPS-enabled cell phones to track the travel patterns of adolescents. *International journal of health geographics*, 7 (1), 22.

Wiehe, S.E., Hoch, S.C., Liu, G.C., Carroll, A.E., Wilson, J.S., and Fortenberry, J.D., 2008. Adolescent Travel Patterns: Pilot Data Indicating Distance from Home Varies by Time of Day and Day of Week. *Journal of Adolescent Health*, 42 (4), 418–420.

Williams, J.C., 2002. A Zero-One Programming Model for Contiguous Land Acquisition. *Geographical Analysis*, 34 (4), 330–349.

Wolf, J., Oliveira, M., and Thompson, M., 2003. Impact of underreporting on mileage and travel time estimates: Results from global positioning system-enhanced household travel survey. *Transportation Research Record: Journal of the Transportation Research Board*, 1854 (1), 189–198.

Xu, Y., Wen, M., and Wang, F., 2015. Multilevel built environment features and individual odds of overweight and obesity in Utah. *Applied Geography*, 60, 197–203.

Xue, Y., Zimmerman, M.A., and Caldwell, C.H., 2007. Neighborhood residence and cigarette smoking among urban youths: the protective role of prosocial activities. *American journal of public health*, 97 (10), 1865–1872.

Yeates, M., 1963. Hinterland delimitation: a distance minimizing approach. *The Professional Geographer*, 15 (6), 7–10.

Zenk, S.N., Schulz, A.J., Matthews, S.A., Odoms-Young, A., Wilbur, J.E., Wegrzyn, L., Gibbs, K., Braunschweig, C., and Stokes, C., 2011. Activity space environment and dietary and physical activity behaviors: A pilot study. *Health and Place*, 17 (5), 1150–1161.

Zhang, L. and Zhou, H.Y., 2004. Research on public establishment location selection based on the voronoi diagram in GIS. *Computer Engineering and Applications*, 9 (223), e227.

Zhang, M. and Kukadia, N., 2005. Metrics of urban form and the modifiable areal unit problem. *Transportation Research Record: Journal of the Transportation Research Board*, 1902 (1), 71–79.

Zhao, P., Kwan, M.P., and Zhou, S., 2018. The uncertain geographic context problem in the analysis of the relationships between obesity and the built environment in Guangzhou. *International Journal of Environmental Research and Public Health*, 15 (2), 1–20.

Zhu, H., Yan, H., and Li, Y., 2008. An optimization method for the layout of public service facilities based on Voronoi diagrams. *Science of Surveying and Mapping*, 2008 (2), 29.

Zick, C.D., Smith, K.R., Fan, J.X., Brown, B.B., Yamada, I., and Kowaleski-Jones, L., 2009. Running to the Store? The relationship between neighborhood environments and the risk of obesity. *Social Science and Medicine*, 69 (10), 1493–1500.

Zoltners, A.A. and Sinha, P., 1983. Sales Territory Alignment: A Review and Model. *Management Science*, 29 (11), 1237–1256.

Zook, M., 2015. Small Stories in Big Data : Gaining Insights From Large Spatial Point Pattern Datasets. *Cityscape: A Journal of Policy Development and Research*, 17 (1), 151–160.